

AN INVESTIGATION OF MULTI-ATTRIBUTE UTILITY  
TECHNOLOGY (MAUT) AS AN EVALUATION METHOD IN AN  
ORGANIZATIONAL TRAINING ENVIRONMENT

by

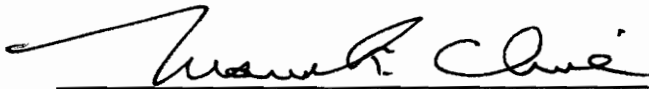
John P. Milatzo

Dissertation submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of  
DOCTOR OF PHILOSOPHY

in

Educational Research and Evaluation

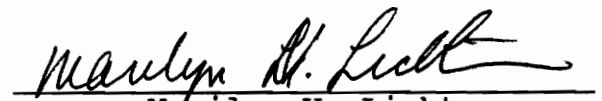
APPROVED:

  
M. G. Cline, Chairman

  
Michael K. Badawy

  
Albert Wiswell

  
Ron McKeen

  
Marilyn V. Lichtman

May, 1993

Blacksburg, Virginia

c.2

LD  
5655  
V856  
1993  
M553  
c.2

AN INVESTIGATION OF MULTI-ATTRIBUTE UTILITY  
TECHNOLOGY (MAUT) AS AN EVALUATION METHOD IN AN  
ORGANIZATIONAL TRAINING ENVIRONMENT

by

John P. Milatzo

Dr. M. G. Cline, College of Education

(ABSTRACT)

Multi-Attribute Utility Technology (MAUT) was investigated as an alternative organizational training evaluation method. Research questions focused on what problems emerge in involving stakeholders in training evaluation, what technical problems emerge in the application of the MAUT steps, and what can be learned about the credibility of MAUT results and their impact on decision making.

Five employee training courses offered by a city government training unit were evaluated. Two groups of stakeholders developed desired course attributes and indicators using MAUT. Survey data collected from students and their immediate supervisors provided indicator measurements. Location measures and utilities were provided to stakeholders and students. Stakeholders then evaluated MAUT from the participant perspective.

Positive and negative findings had implications for the effective use of MAUT. Positive findings were that stakeholders valued the group process and opportunity to have input to the evaluation; surveys were useful for measuring attribute indicators; the development of attributes, indicators, weights and location measures presented no technical problem; and, that MAUT results were credible to students and stakeholders. Negative findings were that stakeholder availability was very difficult to obtain; not all stakeholders could deal with the conceptual tasks of MAUT; surveys were not feasible for attribute development; calculation and interpretation of MAUT utilities were seen as too complex by stakeholders; and, the use of MAUT results to influence decisions was hindered by the perceived complexity and the

absence in the study of the primary decision maker for training.

It was concluded that a major strength of MAUT is that it affords the investigator the opportunity to easily tap into the organizational culture and political setting of a program. In addition, the strengths of MAUT as a training evaluation method are highly dependent on inclusion of all relevant stakeholders, a strong commitment of stakeholder time, stakeholder ability to communicate values and priorities, and, the clear and simple reporting of MAUT results. Detailed recommendations for the effective use of MAUT are provided.

Dedicated to my loving wife Marsha. For your endless support and infinite sacrifice, that I might pursue this long standing personal goal. Thank you for making this possible.

## Acknowledgements

I would like to acknowledge several individuals whose support made this project and achievement possible. First, appreciation and thanks to my advisor and mentor, Dr. M. G. Cline. Dr. Cline was a source of much encouragement and guidance. He was an inspiration and motivator at key points along the way when I had doubts.

I would also like to acknowledge the support of two present and former managers. First, thanks are owed Mr. Richard D. Lake for encouraging me to pursue this degree and for supporting my course work. Second, thanks go to Ms. Yvonne D. Maguire, my current manager for allowing me to continue pursuing this goal under her watch and for allowing flexible work schedules when needed.

Next, appreciation goes to Ms. Angela Duncan, whose superlative administrative and clerical support were given without error or complaint during times of personal and professional uncertainty and disruption in her own life.

Thanks also to my parents, whose values and support for higher education inspired an otherwise average student to reach this high. This is a long way from Wheeler School, folks!

And finally, thanks to my wife Marsha, and daughters Maggie, Jillian, and Gina who have had only a part-time husband and father for much of this project. Your sacrifice is duly acknowledged with much thanks and appreciation. I can now turn my computer over to the children, and pick up shovel, rake and hoe!

**TABLE OF CONTENTS**

	Page
Chapter I - Introduction.....	1
Problem.....	1
Purpose of Research.....	4
Chapter II - Review of Related Literature.....	9
Training Evaluation Models.....	9
Issues in the Evaluation of Training.....	18
Methodological Issues.....	22
Multi-Attribute Utility Technology.....	34
Chapter III - Method.....	49
Overview.....	49
Procedure for this Study.....	53
Chapter IV - Results.....	72
Introduction.....	72
Part I.....	72
Part II.....	109
Stakeholder Availability/Time Commitment.....	109
Difficulty of Activities.....	115
Use of Surveys.....	118
Group Dynamics.....	118
Organizational Politics.....	120
Utilization of Results.....	121
Accuracy/Credibility of Results.....	124
Reaching Consensus.....	125
Miscellaneous Issues.....	126
Chapter V - Summary, Conclusions and Recommendations.....	130
Summary of Purpose.....	130
Summary of Results.....	132
Conclusions.....	136
Recommendations.....	142
References.....	146
Tables.....	74
Table 1 - Potential Standardized Course Attributes Derived from Individual Stakeholder Interviews.....	74
Table 2 - Organizational Values Developed by Executive Groups with Importance Weights.....	76
Table 3 - Executive Group Strategic Value Scores.....	78
Table 4 - Decision Value Tree - Executive Stakeholder Group.....	80
Table 5 - Executive Stakeholder Group Prior Location Measures.....	81
Table 6 - Mid-Management Group Strategic Value Scores.....	82
Table 7 - Decision Value Tree - Mid-Management Stakeholder Group.....	84
Table 8 - Mid-Management Stakeholder Group Prior Location Measures.....	85
Table 9 - Executive Stakeholder Group Indicators.....	87

**Contents (continued)**

Table 10-	Mid-Management Stakeholder Group Indicators.....	88
Table 11-	Survey Response Rate by Course and Survey Type...	89
Table 12-	Executive Group Location Measures & Utilities....	90
Table 13-	Mid-Management Group Location Measures & Utilities.....	91
Table 14-	Executive Group Correlation of Pre and Post- Utility Scores by Course (N=15).....	92
Table 15-	Mid-Management Group Correlation of Pre and Post- Utility Scores by Course (N=12).....	94
Table 16-	Executive Stakeholder Group Utility/Impact Summary.....	95
Table 17-	Mid-Management Stakeholder Group Utility/Impact Summary.....	96
Table 18-	Number of Students (Maximum=7) Responding Positively or Negatively to Interview Questions	98
Table 19-	Summary of Student Reactions to Post-Location Measures, Mid-Management Team.....	100
Table 20-	Revised Student Reactions to Post-Location Measures, Mid-Management Team.....	102
Table 21-	Pearson Product Moment Correlations between Student/Stakeholder Rankings of Course Utility & Total Utility & Total Impact Scores (N=70)...	104
Table 22-	Highlights of Stakeholder Survey Questionnaire Responses.....	105
Table 23-	MAUT Evaluation of MAUT Process in Evaluation of Five Courses (N=4).....	108
Appendices.....		151
Appendix A -	MAUT Instructions by Survey.....	152
Appendix B -	Executive Attributes and Indicators.....	157
Appendix C -	Mid-Management Attributes and Indicators .....	172
Appendix D -	Pre-Training Survey/Supervisor.....	184
Appendix E -	Pre-Training Survey/Student.....	185
Appendix F -	Training Feedback Survey.....	186
Appendix G -	Post-Training Survey/Student.....	188
Appendix H -	Post-Training Survey/Supervisor.....	190
Appendix I -	Student Interview Summaries.....	192
Appendix J -	Final Stakeholder Survey.....	195
Appendix K -	Responses to Final Stakeholder Survey.....	213
Appendix L -	User's Guide to MAUT.....	225
Biographical Vita.....		236



**CHAPTER I**  
**INTRODUCTION**

Problem

The professional literature on organizational training evaluation has frequently noted the relative lack of systematic training evaluation being practiced in business and government organizations (Brinkerhoff, 1987; Latham, 1988; Wexley, 1984; Grider, 1990; Tannenbaum & Yukl, 1992). One reason cited for this is that the level of technical expertise (Brandenburg & Smith, 1986) required often exceeds the capability of organizations. Other reasons are that classical methods of evaluating training do not address the real important evaluation questions; do not provide information to aid in business oriented decision making; are not capable of easily incorporating decision maker values into a quantitative scheme of analysis and evaluation; and, finally, do not enable information to be provided in a timely manner in order to influence decisions (Brinkerhoff, 1987; Hamblin, 1974; Edwards, Guttentag & Snapper, 1975; Edwards & Newman, 1982). The result is that not much evaluation gets done (Brandenburg & Smith, 1986), or, that evaluation efforts fail to influence decisions, and therefore fail to influence the program improvement process.

Multi-Attribute Utility Technology (i.e. MAUT) is an alternative evaluation methodology that has not been thoroughly researched for applicability in the training environment. MAUT is a general analytic method which incorporates decision maker values in a quantitative way to produce evaluation criteria against which to compare objective or subjective assessments of program merit and worth (Edwards, Guttentag, & Snapper, 1975). In MAUT, desired program attributes are defined by decision makers and/or other stakeholders who have a stake in the success of a program.

One of the many strengths of MAUT as a generalized method is its

ability to handle or incorporate experimental or other hard data if they are available. Another strength is MAUT's ability to bring the evaluator in touch with the vital social and political context in which the program exists and in which the evaluation will be conducted.

MAUT was first introduced to the evaluation community in 1975 by Ward Edwards (Edwards et al., 1975) and others in an evaluation handbook paper called A Decision Theoretic Approach to Evaluation Research.

The main theme of that paper was that the nature of evaluation research problems, versus the purpose and requirements of classical experimental and quasi-experimental designs, presents to evaluators and decision makers an inconsistent and often unworkable mix. In short, viewing programs as experiments, a popular notion in the formative years of social and educational program evaluation (Weiss, 1972a; 1972b), presents difficult methodological and practical problems for the evaluator, as well as a communication problem for the decision makers and others for whom the evaluation was conducted.

For the evaluator, the challenge of trying to design and complete the experiment, without contamination and bias, in time to be useful always looms large. For the decision maker and other potential users, reports based on experiments provide information which often does not answer the important questions, and, typically, the conclusions are cast in difficult to understand statistical terms rather than in the language of business. That is, the language of stakes versus odds, economics versus hypothesis testing, and probabilities and percentages versus standard errors and confidence intervals.

Edwards and Newman (1982) outlined six assumptions on which MAUT was based, and which also describe fundamental characteristics of evaluation research problems and the realities of many programs. They are: 1) When possible, evaluations should be comparative, 2) Programs

normally serve multiple constituencies, 3) Programs normally have multiple goals, not all equally important, 4) judgments are inevitably a part of any evaluation, 5) judgments of magnitude are best when made numerically, and 6) evaluations typically are and should be relevant to decisions.

Edwards used these assumptions to define the basis of MAUT, and to say that classical research methods do not easily serve the needs of either evaluators or decision makers. MAUT, on the other hand, comes much closer to meeting the needs and assumptions that normally accompany evaluation problems.

The starting point for MAUT is the notion that evaluations get commissioned in order to provide decision makers with necessary information, and, that rational decisions get made (by individuals or by groups) through a process involving assessing states of the world against criteria which reflect the values of the decision makers about what constitutes "worthwhileness" in the program under evaluation.

To illustrate the applicability of MAUT, a hypothetical example will be offered. Imagine a group of five to eight decision level stakeholders gathered to define evaluation criteria for a program. The opening question by the evaluator might be "When all the data are in, what would you like to be able to say about the program under evaluation?" Sub-questions might be "From your perspective, or in your subjective, value saturated/value contaminated opinion, what is important to consider about this program when considering its value, worthwhileness, etc.?" Another sub-question might be "For a program like this to meet its stated purpose, it must have the following \_\_\_\_\_ . Please fill in the blank with a list of desired course characteristics."

The stakeholders would then attempt to reach consensus on eight to fifteen criteria. Then, they would rank the criteria in order of importance, assign relative weights, and normalize the weights so they sum to 100.

This describes the first essential steps of a MAUT evaluation as outlined by Edwards and Newman (1982). The total list of seven MAUT steps is provided in the review of the literature chapter (Chapter II). They are not described here to avoid redundancy. However, there has been little documentation in the literature about the use of MAUT for training evaluation or of the success or failure of MAUT to satisfy the information and evaluation needs of actual training departments.

#### Purpose of the Research

The purpose of the present study is to examine and evaluate the MAUT process by applying it in an actual training environment, identifying its strengths and weaknesses, and then making recommendation as to its applicability and the preferred method(s) of applying its various steps.

Advocates of MAUT contend that the utility values generated in MAUT, together with the individual attribute utility values, plus the unique way of arriving at those values--beginning with the identification of relevant stakeholders--will lead to superior evaluations. Superior evaluations are defined in terms of having an impact on the decisions that drive the evaluation, or on program improvement efforts.

Thus, according to Edwards and Newman (1982), the advantages of MAUT which set it apart from more classical/traditional evaluation methods are:

- "1) It closes the gap between purely intuitive and judgmental evaluations and the more quantitative kind

- 2) It makes the coexistence of judgment and objective measurement within the same evaluation easy and natural
- 3) It allows for the simultaneous evaluation of both process and outcome goals
- 4) The use of subjective inputs can greatly shorten the time required for an evaluation to be carried out." (Edwards and Newman, 1982, p. 9).

Also, according to Edwards and Newman (1982), a MAUT evaluation can be carried out from original definition of the problem to preparation of the evaluation report in as little as a week of concentrated effort.

The inputs to an abbreviated, one week MAUT evaluation will obviously be almost entirely subjective. But the MAUT technique at least produces an audit trail such that the skeptic can substitute other judgments for those that seem doubtful, and can then examine systematically what the consequences for that evaluation are. The proper comparison to consider, said Edwards, is what happens when time constraints on a decision make haste very necessary? MAUT does provide an orderly basis for decision, and as such, produces both an audit trail, and a more rational (and hopefully better) decision.

The literature on MAUT for educational or organizational training evaluation is scant. The few studies investigating the application of MAUT in educational or training program evaluation focus on comparing the results from a MAUT evaluation with those from an experimental or quasi-experimental design (Hawthorne, 1985; Vernon, 1985), and on the conclusions drawn about the training. These studies failed to describe in detail the process of MAUT or the potential administrative problems encountered during implementation in a training setting. Thus, the present study investigates the administration and outcomes of MAUT, and seeks to obtain insight into its characteristics, strengths and weaknesses in the following areas:

1. Stakeholder Availability/Time Commitment
2. Difficulty of Activities
3. Use of Surveys
4. Group Dynamics
5. Organizational Politics
6. Utilization of Results
7. Accuracy/Credibility of Results
8. Reaching Consensus

These areas represent categories of primarily administrative concerns and issues which the present investigator views as one way or another relevant to the implementation of MAUT in training organizations, and about which very little is written in the MAUT literature.

Evaluation is important to the professional training evaluation community because evaluation ensures program quality (Brinkerhoff, 1987). The investigation of MAUT as an alternative method addresses an apparent need for quality training evaluation in organizations (Wexley, 1984; Latham, 1988).

The relatively expensive cost of training versus other change strategies justifies the cost of evaluation. However, the cost of experimental and quasi-experimental designs for impact evaluation, both in time and in obtaining the required expertise, plus the additional problems associated with these classical approaches (see Chapter II), suggest that other evaluation models such as MAUT are worth investigating as possible alternatives.

It is the thesis of this study that if MAUT could be effective in an actual training environment, it would provide training evaluators and administrators with a more useful evaluation tool which could lead to greater use and practice of training evaluation in general. But first, more needed to be learned about MAUT in an actual training environment.

According to the literature on MAUT, certain other benefits can be realized in the application of MAUT to an evaluation problem. Some of these benefits are directly stated, while others can be inferred.

The following expectations/benefits were extracted or inferred by the author from the literature: 1) MAUT evaluations will make a difference in the decisions that get made about the programs evaluated; 2) Decision makers will view the MAUT technique/approach as appropriate for assessing the strengths and weaknesses of the programs in relevant dimensions of value; 3) A single decision maker is rarely the case in programs; therefore, MAUT will be effective at explicating, comparing, possibly reconciling, and ultimately aggregating the values of groups in conflict about the desired value attributes of programs; 4) Decision makers will find the relatively simple statistics, the large amount of detail, and the quantitative data relevant to their values; and, they will find the reports, which could get thick and heavy (yet thorough), useful and appealing; 5) The decision makers will find the method easy to communicate and useful in situations where time is short and decision makers are numerous and busy; and, 6) The social psychology of small groups and group process effects will be evident in the required stakeholder group activities, such as the development and weighting of MAUT attributes. The present study investigates the MAUT process in the training environment guided by the previous eight (8) focus areas, and by the above list of expectations gleaned from the literature.

The purpose of this research is to examine how MAUT operates as an evaluation process when implemented in a specific setting by an evaluator in a training evaluation context with a specific client. The goal is to generate insights about the operational aspects in MAUT, and to inform potential users of MAUT. The outcomes of this study will be materials and suggestions that might serve usefully in a manual designed

to train MAUT users in the detailed step by step procedures of a MAUT evaluation.

Note well, that these procedures are the operational and interpersonal procedures in carrying out the MAUT technique, and are not limited to the technical aspects that have already been described in the literature (Vernon, 1985; Edwards & Newman, 1982; Edwards et al., 1975; Park, 1979; Hawthorne, 1985; Einhorn & McCoach, 1977; Hogarth, 1979; Larson & Kaplan, 1981; Brumet, 1976; and Baron, 1977). This position has been well stated by Dyer, Fishburn, Steuer, Wallenius, and Zionts (1992) in a recent paper on the future of multi-attribute utility theory: "The object is to identify exciting directions and promising areas for future research. Simple, understandable and usable approaches for solving MCDM and MAUT problems are still needed. Such approaches will undoubtedly be built around decision support systems. The decision support systems that survive and are widely used will have to be user friendly and have other good qualities. Software that has the attributes of easy to use spreadsheets is most desirable." (Dyer et al. 1992, p. 1).



**CHAPTER II**  
**REVIEW OF RELATED LITERATURE**

In order to set the research problem in the appropriate context, and to present the MAUT model in contrast to traditional models and methods, the literature review will be organized around four topics, each topic comprising a section of the chapter. The four topics are:

- 1) What are the most common theoretical models for training evaluation?
- 2) What are the problems in the implementation of evaluations of training programs?
- 3) What methodological issues, process, and implementation issues are relevant to the selection of a training evaluation plan?
- 4) Issues in the implementation and efficacy of MAUT.

1) Training Evaluation Models

The first question to ask is what is training evaluation and where does it fit in the educational research and evaluation field? Training evaluation in the organizational context is much like program evaluation in the educational or social program areas. Brandenburg and Smith (1986) describe training program evaluation as a subset of educational and social program evaluation. Brinkerhoff (1987) says that training evaluation encompasses activities designed to ascertain the merit and worth of training. The merit of training pertains to its effectiveness as a learning vehicle for the trainees, and worth pertains to its effectiveness in helping to achieve productivity, morale, or career development goals of the individual and organization.

In addition to the merit and worth dimensions of training, Brinkerhoff (1987) wrote that training evaluation includes activities designed to determine whether a course is, in fact, needed (at the front end) and whether a course contributes to organizational outcomes or

results, such as lower absenteeism, turnover, or grievances, etc. (at the back end).

To better understand training evaluation models, one needs to understand the notion of the training cycle. The training cycle is based upon the concept of a system or a processing system, and underlies many of the evaluation models developed for training. The systems notion of input-throughput-output is expressed in training and evaluation models as the training cycle. In the training cycle, training is a process that begins with needs assessment and continues through program design, development, implementation, and evaluation phases. The evaluation phase is a feedback phase. Since the corrective nature of the evaluation phase has direct implications for modifying the original program, it is frequently not viewed as the last phase of a five-stage event, but rather as the corrective phase which is relevant to all phases in the process in an ongoing cycle of intervention, feedback, and improvement.

With the above background, training evaluation models will be discussed. One of the frequently cited models since it was introduced in 1959 is the Kirkpatrick model (Tannenbaum & Yukl, 1992). Kirkpatrick described a four-phase evaluation model (Kirkpatrick, 1959; 1960). The four levels of evaluation are referred to as "criterion" levels. They are: (1) Reactions, (2) Learning, (3) Behavior, and (4) Results. Effective training is defined as training that brings positive student reactions, measurable learning gains, changes in job behavior, and positive organizational results. Thus, in Kirkpatrick's scheme the complete evaluation would involve the systematic collection data from all four levels. In actual practice, Kirkpatrick (1978) recommends quantifying student reactions through an end-of-class survey using objective (e.g. Likert-type items) items whenever possible. To measure

learning, he recommends pre-post testing and, by implication, the use of "t" tests comparing pre- and post-mean test scores. To measure behavior on-the-job, he recommends also a pre-post training job behavior measurement instrument of some type. He also recommends the use of control groups, whenever practical, to measure differences between trained and untrained students. However, he does caution that this requires much expertise and time to implement. As a less reliable alternative, Kirkpatrick recommends specifying desired behavior changes in advance of the program, and then preparing interview questions to determine after the course whether students have made any of those behavior changes.

As for results, Kirkpatrick again recommends pre-post statistical measures and control groups where practical. But, he does caution that it is often impossible to "prove" the effects of training or that training benefits outweigh the costs. Frequently, he says, we must settle for "plausible evidence" that the desired results have been obtained (Kirkpatrick, 1977).

The Training Systems Model was developed by Rummler and Brethower (1979). This model views training as a processing system which receives inputs (students), puts the inputs through a planned process called training, then converts the inputs into outputs in the form of better trained/educated resources. Effective training is training that adapts and adjusts to feedback obtained from the receiving system, i.e., the work environment that receives the trainee after training. Against the framework of the general systems model, Rummler and Brethower (1979) recommend an evaluation matrix for viewing evaluation alternatives. Down the left side of the matrix they employ four levels of evaluation "interest": Reactions, Learning, Use on-the-job, and Organizational Impact. Across the top they have five column headings asking: 1) what

might be measured? 2) measurement dimensions; 3) what to look at; 4) alternative data gathering methods; and, 5) evaluation criteria. This framework, plus their systems model, is offered to help evaluators discover the key evaluation questions and measurement variables that will help solve his particular evaluation problem/situation.

As for methods, Rummier and Brethower recommend that evaluators be as rigorous as they can in the "real world." If control groups are not possible, they should consider ABA or multiple baseline time series designs. Although they recognize the limitations of real world "action research," their orientation seems clearly experimental rather than decision oriented. That is, when it comes to whether evaluation is a truth driven versus an improvement or decision driven exercise, Rummier and Brethower (1979) seem to hold to the view that evaluation is a truth-seeking model.

Brinkerhoff (1987) developed a Six-Stage model which imbeds evaluation in each six phases that parallel the training cycle: Analysis (i.e. needs assessment, problem definition, etc.), Design, Development, Delivery, On-The-Job Behavior, and Organizational Results. Evaluation questions unique to each phase guide the collection of relevant information. Brinkerhoff, in contrast to Rummier and Brethower, is much more pragmatic in his orientation. Where Rummier and Brethower would argue for rigor in design to measure training effects, Brinkerhoff (1981) would argue for amassing plausible evidence that links the training context to organizational goals and priorities, and then ties student behavior after training to relevant results linked, as well, to organizational goals.

Nadler (1982) developed the Critical Events Model (CEM) which specifies evaluation as an on-going and continuous process applied to each phase of training design and development process. Specific steps

in the model are (1) Identify Needs, (2) Specify Job Performance, (3) Identify Learner Needs, (4) Determine Objectives, (5) Build Curriculum, (6) Select Instructional Strategies, (7) Obtain Instructional Resources, and 8) Conduct Training. Evaluation questions at each step guide data collection. The focus is mostly on formative evaluation designed to improve training while in the development stages.

The Instructional Systems Design (ISD) Model (Rosenberg, 1982) also directs that evaluation information be collected throughout the design and development process. The design and development process is comprised of four phases--analysis, design, development, and implementation. Evaluation is the systematic process of verifying that the program development is proceeding as expected at each phase.

Hamblin (1974) specifies a five-stage training and evaluation model. In this model, training evaluation is a quality control process that has relevance at four key points along the training continuum: Student Reactions, Learning Behavior, Organizational Results, and Ultimate Value. Hamblin has essentially taken Kirkpatrick's four-stage model and broken stage four out into two stages -- Organizational Results pertains to "measuring the effects of trainees' job behavior changes in the functioning of the organization in which they work." (p. 22). To evaluate at this level, it is necessary to specify organizational goals and objectives at the office or department level, and these will vary depending on the organization. In contrast, the Ultimate Value level pertains evaluation aimed at the ultimate criteria by which the organization judges its success or failure. In industrial organizations in a capitalist society, these criteria are likely to be economic. Hamblin also calls this the cost-efficiency level. The question usually asked is, "Are the financial benefits resulting from the training greater than the financial cost of the training?"

Other less well-known models also appear in the literature. Elsbree and Howe (1977) presented a three-part evaluation model aimed at satisfying the information needs of the evaluation clients, i.e., the people making decisions about the training. They specify that the evaluation process should consist of a Focus phase, a Plan phase, and an Implement phase. In the Focus phase, the evaluator works with the client to establish the general scope and objectives for the evaluation effort. Next, a Plan is drawn up and shared with the client for input and approval. Finally, the Implement phase is the execution of the Plan. Data is collected, interpreted, and provided to the client with recommended action. The theme of this model is that if the client is not on-board with the evaluation objectives and scope, and if the evaluator is not sensitive to client needs, financial constraints, and the political issues, then the evaluation is headed toward failure as a program improvement effort.

Bakken and Bernstein (1982) specified a SET (Systematic Evaluation of Training) model for the evaluation of training. The SET approach considers specification of the decision makers as the initial step. At this step, it is important to consider management decision makers in addition to trainers or training function decision makers. The authors point out that decisions made by Management are fundamentally different from those made within the training function, thereby calling for different information needs. Step 2 is clarification of the training goals and objectives. Four categories of training goals are specified: 1) Personal Growth, 2) Acquire Knowledge, 3) Improve Performance, and 4) Improve Organization. The authors cite these as modifications of Hamblin's (1974) five categories of evaluation. Bakken and Bernstein develop a matrix similar to Rummeler and Brethower (1979), with the four categories of training goals on the vertical axis, and Kirkpatrick's

four criteria levels on the horizontal axis. The purpose is to allow the evaluator to select a measurable training outcome that is appropriate to the particular objective of training and to the needs of various decision makers. Once it is determined what will be measured or what data will be collected, however, the authors then direct the evaluator to standard research concepts such as validity, reliability, pre-post testing, control groups, and time series designs. Thus, the model is useful for specifying what will be measured, but then turns the evaluator over to the common data collection and analysis issues of the social and behavior sciences. These, in turn, are rooted in classical statistics and randomized control group experimental design.

Snyder, Raben, and Farr (1980) introduced a model similar to Stufflebeam's (1971) CIPP model. The starting point for their model is the notion that Human Resource Development (HRD) programs are often open-ended systems with no effort made to evaluate their effects. Such a lack of comprehensive evaluation practices, they say, seems grossly out of line with current organizational theory and the availability of various behavioral sciences-based evaluation methods (note here also the influence of traditional "classical" behavior sciences orientation among the authors). The authors also note that the concept of evaluation is a specific case of the concept of experimentation. In other words, they see evaluation as establishing relationships and degrees of change in outcomes based on changes or manipulations in antecedent (independent) variables. Moreover, two major research objectives of any program evaluation must be to: 1) identify the nature and range of changes produced by a program, and 2) to show how and why the program has created these effects. The interpretation of the effects produced by a change program as desirable or "effective" becomes, for Snyder et al. (1980) a "separate but related issue that must ultimately be addressed"

(p. 433). (As we shall see later, this experimental view of evaluation runs counter to the underlying basis of MAUT). In any event, Snyder et al. (1980) go on to specify their model of evaluation as a feedback cycle made up of Context, Input, Process, Summative, and Prognosticative Evaluation (CIPSP). Context evaluation consists of an assessment of the operating context of the organization involved. Input evaluation is to assess available organizational resources and operating policies to establish the feasibility of various programs intended to ameliorate the problems identified in the Context evaluation. Process evaluation involves monitoring the processes and procedures of the program while it is actually functioning and feeding data back to organization members and decision makers. Summative evaluation concerns the effectiveness and efficiency of the program as an organizational change agent, and Prognosticative evaluation involves generalizing from the program under investigation to the likelihood that a program of the same type would have similar or different effects in other parts of the organization.

In addition to the above models, educational and social program evaluation models have influenced the thinking and practice of training evaluation and, therefore, will be mentioned briefly.

Tyler (1950) introduced the notion that education and training programs should be designed around, and evaluated against, a specific set of behavioral objectives. Evaluation should collect information to determine whether the specific objectives of the course have been met.

Stake's Countenance Model (Stake, 1967) emphasized the importance of judgement in the evaluation process. In this model, two sets of information need to be collected about the evaluated object: descriptive and judgmental. The descriptive set should focus on intents and observations regarding antecedents (prior conditions that may affect outcomes), transactions (the process of implementation), and outcomes.



The judgmental set is comprised of standards and judgments regarding the same antecedents, transactions, and outcomes.

The CIPP (Context, Input, Process, and Product) model (Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman, & Provus, 1971) considers four things about a program under evaluation: (1) its goals; (2) its design; (3) its process of implementation; and, (4) its outcomes. Evaluation is an assessment of: (a) the merit of goals; (b) the quality of plans; (c) the extent to which plans are carried out; and, (d) the worth of outcomes.

The Goal-Free Model (Scriven, 1967) directs the evaluator to collect data about program processes and products without having prior knowledge about the goals and objectives of the program. Prior knowledge of goals is thought to unnecessarily bias the evaluator and interfere with accurate judgements of value and worth.

Guba and Lincoln (1981) suggest five kinds of information for program evaluation: (a) descriptive information about the object evaluated, its setting, and surrounding conditions; (b) information responsive to concerns of relevant audiences; (c) information about relevant issues; (d) information about values; and, (e) information about standards relevant to merit and worth assessments.

In summary, a few common threads among the training evaluation models can be identified. First, training is commonly viewed as a cyclical and sequential process made up of stages. Second, evaluation is viewed as a feedback and corrective process relevant to each phase. Third, evaluation is not viewed as a single variable, one-time measurement activity. Rather, evaluation is a complex and multi-dimensional process of a different order than the one variable hypothesis testing model of classical experimental design and research. (MAUT, in comparison to these models, is particularly well-designed to

accommodate multi-dimensional criteria). Finally, several models say it is important to look both within a course as well as beyond (i.e. after) a course to address both questions of training merit and training worth.

## 2) Issues in the Evaluation of Training

The question of frequency or amount of training evaluation being conducted in organizations has been addressed by several writers. The conclusions have varied by author and study. However, the most predominant theme has been that, for all the virtues, common sense, and good business justifications available for doing evaluation, the practice of evaluation in training organizations is lacking in rigor and quality if not in frequency of occurrence. The reasons given for this are usually political, economic, and technical (Geroy and Wright, 1988).

The political issue is one of pride and ownership in the training curriculum by the Training Department, plus a mistrust of an independent evaluation process. On the one hand, program directors and administrators often have a large ego investment in their programs. If the program isn't "broke", they don't want anyone to "fix-it," period. The notion that their program would be examined by evaluation "experts" is particularly troublesome to those whose programs have provided their sponsors with much positive recognition, power, and status.

The economic issue concerns the cost of conducting evaluation. It is often the case that program directors are under pressure to lower their training budgets in deference to necessary funding for operations. When organizations begin to tighten their belts, training programs are often considered more expendable than other programs more directly involved with the organization's mission and primary products. In that scenario, training departments would be reluctant to spend dollars evaluating programs when all available funds are needed just to keep the

[favorite] programs running in the first place.

The technical issue has to do with the mechanics of designing and conducting evaluation. To the extent that evaluation is an academic discipline, and one that has techniques and standards normally conveyed through graduate study in universities, organizations are faced with fighting for personnel dollars to hire trained evaluators in addition to trained trainers. Again, when funding is hard to come by, the priority would go to training personnel just to keep programs operating. (Nevermind that timely evaluation could have saved an otherwise troubled program from its own demise!)

In terms of the literature, however, since 1966 writers have lamented the lack of quality, comprehensive evaluation within training functions in business and government (Blumenfeld, 1966; Dubin, Mezack, & Neidig, 1974; Goldstein, 1974; Hinrichs, 1976; Goldstein, 1980; Clement, 1981; Zenger and Hargis, 1982; Goldstein, 1986; Wexley, 1984; Brandenburg and Smith, 1986; Latham, 1988). The claim is that training evaluation has lagged behind the growth in training activity in organizations.

Although the amount of unpublished evaluation being conducted in organizations is difficult to estimate (Hawthorne, 1985), some researchers have attempted to answer the questions of amount of and type of evaluation being done by surveying training organizations.

Catalanello and Kirkpatrick (1968) surveyed 154 firms to determine the amount of formal evaluation designs being used for training evaluation. Out of 110 responses, they found one firm that used experimental and control groups for evaluation.

Campbell, Dunnette, Lawler, and Weick (1970) reviewed 73 studies and found that 45 used before and after control group designs, and the remaining 28 used only before and after measures with no control groups.

Ball and Anderson (1975) surveyed several organizations and found a preponderance of one-group observational studies being used. However, they did observe that 81% of the respondents indicated that evaluation did, in fact, influence decision making in their organizations.

DeMeuse and Liebowitz (1981) reviewed 36 studies and found 21 used pretesting and 18 used multiple group designs. But overall, 20 or 36 studies were called "pre-experimental" in terms of the Campbell and Stanley (1966).

Clement (1981) reviewed 26 studies after 1970 and found that 58% used control groups (29% pre-1970); 19% compared the relative effectiveness of multiple methods for achieving the same training objectives (5% pre-1970); 8% measured individual differences (11% pre-1970); and 4% measured the environmental effects of the organization on the effectiveness of training (7% pre-1970). Clement concluded that not much progress had been made since 1970 on the quality and rigor of methods used to evaluate management training when one considers all the categories in his study.

Brandenburg (1982) surveyed 22 members of the Chicago Sales Training Association (a subgroup of American Society for Training and Development) and 28 members of the Training in Business and Industry Special Interest Group of the American Educational Research Association. The purpose was to find out how training evaluation was being practiced by asking practitioners of the training function in industry.

The results indicated that the highest ranked evaluation purpose was to improve training programs. The second highest was to provide feedback and information to decision makers. In contrast, studying trainee effectiveness after training was ranked eighth out of a possible ten. The most frequently used technique was the student reaction survey--open-ended questions and Likert-type objective items.

Brandenburg's conclusion was that the student reaction "smile sheet" continues to be the most frequent data collection technique, followed by multiple-choice tests to measure learning. In addition, the most frequently cited evaluation purpose was program improvement.

Smith (1984) reviewed 331 studies and found that 41% used pre-testing, 41% used control groups or equivalent, but only 17% used randomly selected or matched samples.

Grider and Capps (1990) reported a study conducted among 1200 members of the American Society for Training and Development. Respondents were asked to indicate the most effective techniques for training evaluation, and indicate whether and why they used those techniques. The results indicated a large discrepancy between preferred evaluation techniques and actual use of them. Major constraints cited for training evaluation were budget, time, and labor required.

To summarize or draw conclusions from these survey studies is difficult. The main reason is that they all varied in scope, the nature and type of organization surveyed, the types of training investigated, and the sample size. However, one common point of interest in the studies was whether experimental and control groups were used in the evaluation. Only Brandenburg (1982) appeared to express any obvious interest in finding out what organizations thought the purpose of training evaluation was, and only Ball and Anderson (1975) showed interest in determining whether evaluation was influencing decision making. The interest in whether control groups were used indicates, in the writer's view, that those conducting these surveys were of the view that quality evaluation by necessity includes some form of experimental design. This predominant theme may have something to do with who is conducting the survey in terms of their status as academicians versus applied practitioners of evaluation. According to the short biographies

appearing in the articles, only Smith (1984) was not affiliated with a university. As for those who have reviewed the field for such journals as the Annual Review of Psychology, those writers also tend to be from academia and are also those who have lamented the lack of experimentation as well as lack of evaluation in general.

This issue notwithstanding, some conclusions across these sources will be ventured. First, student reactions (i.e. the "smile" sheet) are the most frequent form of evaluation. Second, student behavior after training is the least frequent source of evaluation information. Third, the use of experimental designs or control groups exists in the literature, but not in overwhelming degrees. Finally, this writer is of the belief that the amount of systematic evaluation activity in business and government training units has a long way to go to catch up with the amount of training being conducted. Nothing in this review of the status of training evaluation was uncovered to run counter to this belief.

### 3) Methodological Issues

The selection of a model to conduct evaluation, and even the development of a model, is not unrelated to one's view of the role of evaluation. The role of evaluation, in turn, is not unrelated to the data collection, analysis, and reporting methods one employs. If one views the role of evaluation as research to establish causality, one would be inclined to select experimental designs or related methods. Thus, the role of evaluation and methodological issues merit discussion.

According to Brinkerhoff (1987), evaluation is necessary because training programs, like the humans that create them, are fraught with error and imprecision (Brinkerhoff, 1987). The role of evaluation is to keep programs on course through systematic feedback in a constant cycle

of action-feedback-correction.

Brandenburg and Smith (1986) defined evaluation as the process of rendering a judgement of some entity (e.g. a course or program) on some "dimension of value to the client" (p.2). (The importance of stakeholder values in evaluation will be seen later in MAUT.) Goldstein (1974) defined evaluation as the process of systematically collecting the descriptive and judgmental information necessary to make effective training and educational decisions. (The use of the terms "judgmental" or "subjective" will also be seen in MAUT.)

Hamblin (1974) defined evaluation as the assessment of value or worth. Consistent with the evaluation cycle concept, Hamblin viewed evaluation as an ongoing, self-correcting feedback loop necessary for program improvement.

Hamblin, Brinkerhoff (1987), and others view training as having an underlying logic about what is expected to happen as a result of training. The evaluator's role is to define the underlying program logic and to collect relevant data to gauge the extent to which the expected chain of events did, in fact, occur. The purpose of evaluation is to discover what went wrong (or right), i.e. which links in the chain are broke, in order to make program improvements.

Nadler (1982) described evaluation as not a single activity but an ongoing process to aide in the improvement of programs. Nadler (1982), Brinkerhoff (1987), Goldstein (1986) and others see evaluation as a feedback process which is relevant to all stages of the [needs] analysis, design, development, and implementation of programs.

Scriven (1967) wrote that evaluation can serve two primary functions, the "formative" and the summative." Formative evaluation is used for the improvement and development of an ongoing activity. Summative evaluation is used for accountability, certification, or

selection. In the training context, formative evaluation usually refers to assessing the program merit, that is, "Are we teaching things right?" Summative evaluation refers to the program worth, which is an after the fact results-oriented question "Are we teaching the right things?" (Brinkerhoff, 1987).

Another function of evaluation is the psychological or socio-political function (Nevo, 1986). This refers to evaluation being used to increase awareness of special activities, motivate desired behavior of those evaluated, or promote public relations.

Evaluation differs from research, and the experimental model has long been recognized as being best suited to research (Weiss, 1972a, 1972b; Rossi, 1972). According to Nadler (1982), training evaluation is concerned with determining whether the instruction was delivered according to plan (i.e. the instructional design); whether students were able to learn and retain from the instruction; whether the instruction transferred to the jobs of the trainees; whether the expected benefits of the instruction were realized [back] on-the-job; whether the organization as a whole changed as a result of the training; and, whether the problem identified in the [needs] analysis was solved. In addition, evaluation is often measured against the learning objectives stated prior to the program, and the focus is on decisions that have to be made about the condition and future of the program.

In contrast, research is a truth-seeking, hypothesis testing endeavor focused more on how and why the learning took place. The goal is understanding phenomena, with an eye toward identifying generalizable laws and principles which go beyond the particular learning situation under study.

Research often involves theory building, theory testing, and verifying predictions made from theories and prior research.



Evaluation, on the other hand, uses the actual learning situation to determine what should be done next toward program improvement, program funding, etc. Research frequently requires carefully controlled experiments to determine casual relationships between variables. Although evaluation can often make use of findings from controlled experiments, it is infrequent that situations require such level of rigor and control, or to establish casual relations among all relevant variables. If the goal is to keep on target, then discovery of when you are off target is critical, and tests of hypotheses are less relevant.

Hamblin (1974) also distinguished evaluation from research. For Hamblin, the purpose of evaluation is to provide useful information, not proof in the "scientific" sense. Hamblin argued for an evaluation strategy called "Discovery" as opposed to "Scientific". In the Discovery mode, the activities are flexible; the evaluation and the training are intermingled (i.e. imbedded); and the goal is to create an ongoing and integrated training, feedback, and evaluation system. In the Scientific mode, the activities are more rigid; the evaluation is kept separate from the training; and, the purpose is to "prove" the training is effective.

The techniques and data collection methods of evaluation and research often appear similar, and this has contributed to confusion among the terms (Nadler, 1982). Both are concerned with the collection of accurate information, and both often use statistics to gather and report data. However, research, by nature, is concerned with adding to a theoretical system based on knowledge and truth. Thus, research tends to use more rigidly controlled data collection procedures and experimental designs in order to identify effects to the exclusion of competing alternative explanations of certain phenomena. The randomized controlled experiment is the preferred method for controlling for the

common threats to internal and external validity (Campbell and Stanley, 1966).

The sources cited in the review of training evaluation models, together with the references on the role of evaluation indicate a strong need for a pragmatic research strategy rather than a truth-seeking paradigm. A primary thesis of the present study is that traditional research methods, including randomized control group experimental designs, are less useful for training evaluation for both philosophical, technical, and practical reasons. This idea will be supported below.

Geroy and Wright (1988) provide arguments for the pragmatic over the truth-seeking strategy for training evaluation. They contend that training managers, training and development practitioners, and education professionals are confronted regularly with program related problems requiring practical solutions. Frequently, decisions are based on "gut" feelings, or on experience gained from previous solutions which were used with varying degrees of success in similar situations, rather than on relevant data derived from systematic research or evaluation activities. If the "gut" feelings are right, heroes are made. If the gut feelings are wrong, disaster could result. Yet, much decision making continues to be based on the informal data base of the "gut" feeling.

Several factors account for the perpetuation of this method of decision making (Geroy & Wright, 1988). First, in many work environments, there is short time available to gather data in support of the decision making process. In other instances, practitioners have little formal preparation in research methods or strategies. In addition, many practitioners consider "traditional" research methods difficult to apply, are inappropriate, or do not lend themselves to practical solutions because the control of variables in business

solutions often is difficult.

Despite these and other observations about the lack of applied systematic procedures for problem solving, the perception held by training professionals is that more applied methods are needed to support the decision making function. As Geroy and Wright (1988) note: "What appears to be needed is the development of research methods or strategy that training practitioners consider pragmatic and that can be applied and managed systematically" (p.18).

As will be seen, the MAUT procedure has appeal to those who support this position. Until fairly recently, technical quality and accuracy were the primary concerns of evaluation researchers (Geroy & Wright, 1988), and the experimental method was the preferred evaluation design. Methodological rigor dominated the evaluation decision process rather than pragmatism, and was often the only criterion by which evaluations were judged. Methodological rigor meant experimental design, quantitative data, and detailed statistical analysis.

In order to understand that the experimental tradition is not entirely suited to the pragmatic requirements of program evaluation, the philosophical roots of research and evaluation must be described.

Research has its roots in logical positivism, and evaluation can be traced to pragmatism as philosophical underpinnings (Geroy & Wright, 1988). Logical positivists were pure empiricists who viewed the world as being composed of one set of underlying truths about how things worked and how variables related to one another. Consistent with this view, the goal of the methods of science are to isolate direct effects of independent variables of interest on dependent variables of interest for the purpose of finding out "what caused what." The ultimate goal is to explaining how variables in the "black box" interrelated through the development and testing of theories and hypotheses. The classical

hypothetical-deductive method is well suited for that type of investigation.

Pragmatists relate the meaning of any concept to its practical consequences (Geroy & Wright, 1988), if the pragmatists believe that the whole truth of that concept is known or understood. In addition, pragmatists view the likelihood of data or information retaining truth from one problem to another as very small.

Evaluation, and the pursuit of value and worth as defined in earlier sections, implies an underlying pragmatic, problem solving orientation. Research, with experimental design as its premier model, is positivist in nature, and thus not easily suited to problems requiring practical solutions and assessments of value and worth.

Thus, the consequences of choosing an experimental design for program evaluation in field settings are problematic by philosophical foundation alone. In addition, there are both technical and practical difficulties and barriers which further militate against their use. A review of these problems is in order.

(1) Randomization: The preferred way of dealing with threats to internal and external validity is random assignment to treatment and control groups (Campbell & Stanley, 1966). One problem with randomization in field settings is that the control group by definition ends up being denied training--at least temporarily. Although Cook and Campbell (1975) have offered several creative solutions for randomization aimed at minimizing negative impact on the organization (e.g., temporarily withholding treatment in the control group, taking advantage of opportunities where demand exceeds supply, opportunities where lotteries are acceptable, etc.), organizational resistance to the concept of randomization makes feasibility a barrier administratively and politically, if not technically (Brewster, 1980).

Quasi-experimental designs are a less preferred alternative to full experiments when control groups or randomization is not possible. However, other problems are created. For example, multiple-baseline time series designs (Komaki, 1977) require the organization to tolerate multiple repeated measures of the same population over time, something it may not be willing to do.

(2) Administrative Control of Experiment: A major requirement for experimental and quasi-experimental designs is that sufficient control be maintained over the execution of the experiment itself. In the ideal case, for example, subjects would not drop out through attrition, control groups would actually control by not seeking training elsewhere to solve the immediate problem, subjects selected for training would cooperate and attend the training, subjects in training would not speak to subjects in control group thereby avoiding treatment contamination, and all subjects in the treatment (i.e. training group) would receive the same amount of training in exactly the same manner (Cook & Campbell, 1975); subjects would not get transferred or promoted until the experiment was completed, and line management would cooperate in such things as post-course performance ratings of former trainees (Brandenburg & Smith, 1986). As the above list suggests, it is not difficult to see the practical difficulties of conducting controlled research in a live business environment.

(3) Criterion: Another category of issues involves the reliability and validity of the criterion serving as the dependent variable for the experiment. Goldstein (1986) devotes a whole chapter to the requirements for the criterion in experimental designs. In addition to the two requirements of reliability (i.e., accuracy and consistency) and validity (i.e., relevancy), Goldstein lists two other factors important to the determination of whether a criterion will be

useful. They are: (1) deficiency--the extent to which components are identified in the needs assessment but not in the criterion, and (2) contamination--the existence of extraneous elements in the criterion which result in the measure inaccurately representing the construct identified in the needs assessment.

In addition to these standards for the criterion, other issues can militate against choosing an experimental approach: (1) changing conditions in the work environment which require changes in the criterion (Brandenburg & Smith, 1986); (2) deciding when to measure, i.e., how long after the training should elapse before valid measures of performance are possible (Brandenburg & Smith, 1986; Kirkpatrick, 1977); (3) if post-course ratings from immediate superiors are used, how does one mask the fact that the trainee did in fact attend training? Just knowing this may bias the rater to rate higher than he otherwise would; (4) reaching consensus among stakeholders on what the criterion variable is, could be a problem, as well as the general complexity of measuring job behavior (Brewster, 1980); (5) the existence of vague program goals may render the criterion impossible to specify (Weiss, 1972a, 1972b); and, (6) "worthwhileness" (Rossi & Williams, 1972), a logical criterion for decision makers, is not a unidimensional variable. The implication here is that a multivariate study of sorts is required, but then the technical complexity of the study rises significantly for practitioners.

(4) Unstateable Hypothesis: This issue concerns the fundamental requirement of experiments that they culminate in statistical tests of hypotheses. If hypotheses are to be tested, then they are to be developed from the research questions. But like many qualitative research questions (Bogdan & Biklen, 1982; Guba & Lincoln, 1981; Patton, 1980), many evaluation questions are simply not stateable as hypotheses (Cline, 1988). For example, the following questions are very relevant

in program evaluation, but researchers would be hard pressed to form classical, statistically testable hypotheses from them: (a) Is this program worthwhile? (b) Is the program useful? (c) What do the students prefer? (d) What do the students think? (e) What do the stakeholders value? (f) What is the nature of the treatment? (g) Is the training design efficient?

(5) Values: Experiments cannot easily handle value questions that are so fundamental to the evaluation and decision making process (Edwards et al., 1975; Edwards & Newman, 1982). Decision theorists tell us that decision making is fundamentally a process of determining what is at stake, and what are the odds that various events are likely to occur. It is said that experiments are very good at assessing odds, but are unable to deal with identifying and quantifying stakes (Edwards et al., 1975; Edwards & Newman, 1982). This is related also to the criticism that after the statistical test for significance is made, the experimental researcher is left with having to assess "worthwhileness" (or value) and the decisions which drove the study from the beginning. This will inevitably involve values and subjectivity which, from the start, were treated as a source of error and contamination rather than as an issue that needs to be dealt with in order to make a practical decision.

(6) Narrow Focus: Most single-variable experimental designs focus narrowly on the likelihood that a certain treatment (i.e. training) would bring results of a certain magnitude larger (or smaller) than the control group or alternate treatment group. In the training situation, the hypothesis to be tested would typically be whether the group that had training scored better than a control group on some key dependent variable. The trouble with this for decision makers is that a statistical test of whether a trained group did better than a control

group is but one in a number of possible areas of interest bearing on the decision to take significant action involving a course. In other words, where does it leave the decision maker when it is found that the training had no statistical effect over the control group? This information of itself does not enable the decision maker to develop an improvement strategy for revising the training. Yet, a well-designed experiment to answer the "significant effect" hypothesis has probably cost the decision maker quite a lot of money only to learn that the training had no effect, that it's back to square one and try again. The inherent nature of experimental analysis of variance design and procedure is, therefore, problematic to evaluation done for program improvement.

(7) Complexity of Programs: The sheer complexity of programs often makes the parallel between them and univariate independent treatment variables, such as drugs in a laboratory study, invalid. Yet, the experimental paradigm calls for the independent variable (e.g. a training course) to be homogeneous in nature, uniform and constant across subjects, unchanging over time except when it varies by well-planned, well-controlled design. Technically, if one were to conduct an evaluation with an instructor-lead training program as the "treatment", the best way to ensure purity of treatment would be to have the same instructor give the training in exactly the same way over the duration of the experiment, much like what would be the case with a self-paced programmed text. Yet, the very nature of the dynamics of management training courses often requires that the student mix in any given class must dictate the relative time spent on various topics. This leads to variations in knowledge and learning acquired from class to class. Add to this problem the operational realities that could serve to interfere with the experiment, such as (1) the need for multiple classes running



simultaneously to meet operational needs, (2) program directors who insist on changing and improving the program on the spot as new information and ideas for improvement emerge, and (3) the need to change technical training courses in response to changes in the environmental conditions, one can see more clearly the barriers that program complexity adds to the situation. (Note: Although program complexity can be a problem in any method or design including MAUT, it is a particular problem in experimental designs where specifying program effects and causal attribution are the goal. MAUT, in contrast, was developed to aid in decision making rather than as method to isolate program effects and attribute cause.)

(8) Miscellaneous: Additional problems cited in the literature for conducting field experiments are: costly, time consuming, and complex (Kirkpatrick, 1977); statistical power and complexity of programs (Lipsey, 1988); negative stigma of the "mad scientist" in organizations (Rossi & Williams, 1972); low evaluability of experiments due to measurement unreliability, designs which can't be implemented, and data which can't be analyzed (Rutman, 1980).

In summary, this review has presented the more common training evaluation models, the status of training evaluation, and the primary methodological issues surrounding the use of classical research methods, such as experimentation. In this writer's view, some of the evaluation models imply experimentation, or at least the use of classical statistics, as the primary analytical tool. Other models speak of the need to incorporate decision maker needs and values into the evaluation design. But, none of the current models speak to how one could deal with decision maker values quantitatively. When the methodological issues are added to this situation, the result is an evaluation need lacking a workable analytical framework and paradigm.

In 1975, MAUT methods were introduced in the evaluation literature (Edwards et al., 1975). MAUT was introduced as a quantitatively oriented, pragmatic alternative to experimental designs. The next section presents a definition and description of MAUT.

### Multi-Attribute Utility Technology

a) Implementation of MAUT: Multi-Attribute Utility Technology (MAUT) first appeared in the program evaluation literature in 1975 as an alternative to the experimental method for program evaluation (Edwards et al., 1975).

The main foci of MAUT are the decisions to be made or the questions to be answered about programs, and the values of those who will make those decisions or answer those questions (Vernon, 1985). Values are relevant in MAUT because values are inextricably linked to the criteria decision-makers use to evaluate programs, and evaluation criteria (also called dimensions or attributes in MAUT) are the means and basis by which the decision maker(s), called stakeholders, make their decisions or assess the worth of programs.

Subjective as decision criteria may be, the MAUT process captures, quantifies, and makes explicit the decision criteria that are inexplicably part of the judging, evaluating, and decision making process. As such, MAUT has roots in decision theory and Bayesian statistics (Edwards & Newman, 1982).

MAUT can incorporate any information developed from experimental or quasi-experimental research, if it is available (Edwards et al., 1975). Classical research results are accepted and can be accommodated in a MAUT decision making system, but much consideration is given to the human situation that goes beyond classical experimental methods, and allows the use of data that are relevant and specific to a particular

decision or program evaluation, even though they may be too "subjective" for most advocates of the more classical approaches.

Several sources describe the essentials and roots of MAUT (Vernon, 1985; Edwards & Newman, 1982; Edwards et al., 1975; Park, 1979; Hawthorne, 1985; Einhorn & McCoach, 1977; Hogarth, 1979; Larson & Kaplan, 1981; Brumet, 1976; and Baron, 1977). In the MAUT (i.e., decision-theoretic) model, the role of the evaluator is that of a facilitator for decision making or program evaluation (Baron, 1977). The evaluator collects data and presents it to the decision maker who will make a decision. The most important question for the decision maker is "What data should be collected?" The decision-theoretic model is most useful for answering this question because it is derived from the assumption that people make decisions by evaluating various entities (alternatives) on many relevant value dimensions.

If two programs need to be compared, for example, program effectiveness will be an important dimension. But in the typical classical/experimental design, effectiveness will likely be the only criterion. In the contrasting decision-theoretic model, many additional important factors can be accommodated. The model acknowledges the multi-faceted complexity of decision making and evaluation. It attempts to quantify the decision/evaluation process by isolating the values held by the decision maker/stakeholders and prioritizes them in the way he or she does when making a decision. Data are then collected on the important dimensions, and the data can be used to modify prior expectations into a newly integrated assessment of facts and probabilities about states of the world. Thus, a decision is made using the best information available which, in the real world, is almost always a combination of old and new information.

The essentials of the MAUT method are described as series of developmental and analytical steps (Edwards & Newman, 1982):

1. Identify the objects of evaluation (called entities), or the functions or processes which the evaluation is to perform.
2. Identify the stakeholders--Those people who have an interest or "stake" in what is being evaluated and/or who are important enough so that their interests should be considered.
3. Elicit from stakeholders a common set of the relevant value dimensions or attributes which the program(s) being evaluated must/should hold. These are the "criteria" by which the various stakeholder groups do and will judge the program(s).
4. Elicit from each stakeholder group the relative importance (i.e., weights) of the items comprising the set of attributes or value criteria.
5. Identify, through empirical research, or systematic subjective ratings, how well each program serves each value or attribute criteria. (called establishing location measures)
6. Aggregate the location measures with measures of importance. Essentially, this involves multiplying each location measure times the weight of the criterion or attribute and summing across attributes to get a utility value for each program.
7. Perform a sensitivity analysis by manipulating the numbers to see what effect increasing the value of any attribute or increasing the location measure will have on the utility index.

b) Efficacy of MAUT: The references cited above are primarily theoretical papers which speak to the principles, advantages and mechanics of MAUT. This section reviews recent research literature of the application of multi-attribute techniques applied to concrete situations where multiattribute as a generic descriptor was involved somehow in the research, and focus on those few papers where MAUT as described the sources in the previous section was applied in the education or training environment.

Garvill, Garling, Lindberg, and Montgomery (1992) used a multiattribute set of stimuli to study consumer preferences versus consumer choice behavior in the area of selecting personal residence. That is, research on consumer preferences for residential attributes has shown that cost appears to be more important when people actually choose a residence than when they state their preferences about various attributes of residences. Garvill et al. (1992) attempted to replicate this finding by presenting subjects with a set of residential attributes, having them rate the desirability of them, and then using this information in conjunction with other variables in two experiments designed to test the differential predictability of preference versus choice. The subjects did not develop or weight the attributes, but were simply presented with the rating task.

Alpert, Kamins, and Graham (1992) used a multiattribute model to develop a questionnaire that was administered in an experiment to study retail buyers attitudes toward products which are pioneers in their category versus products which are called "me-too followers." They then used data from the questionnaire to test a causal model for determining why buyers show such an attitudinal preference for pioneer brands. As in the previous reference (Garvill et al., 1992), the researchers did not describe how they developed the multiattribute

research instrument, nor whether they used any of the techniques and concepts Edwards and others described in their papers on MAUT.

Kopalle and Hoffman (1992) used simulation and algebraic analysis to study attribute weight sensitivity in a consumer preference context. What they found was that they could determine whether a particular choice of weights was more or less sensitive to the determination of an overall product quality scale than others. This was basic research on the characteristics of attribute weights in consumer preference research.

Teas and Wong (1992) studied the test/retest reliability of attribute ratings using undergraduate business students who were rating department store attributes. The researchers developed the attributes and produced the questionnaire themselves. The study found that the students were consistent in their ratings over time. Sanbonmatsu and others (Sanbonmatsu, Kordes, & Herr, 1992) studied the conditions under which people attend to the absence of important information and where such a recognition or detection subsequently influenced their judgement or decisions about consumer products. The authors developed a questionnaire containing potentially desirable attributes of ten speed bicycles and had subjects rate attribute importance under two experimental conditions. What they found was that the more expertise people have about a product, the less likely they will fail to recognize important missing information when they are forming judgements about products.

Garre (1992) noted that MAUT is a practical, effective in group decision making, and outlined steps for its use for managers of health care facilities. However, the outline of steps lacked much of the detail on the MAUT process that is the object of the present study.

Dyer, Fishburn, Steuer, Wallenius, and Zionts (1992) reviewed the

history of MAUT and Multiple Criteria Decision Making (MCDM) and other topics believed to be important to management science over the next decade. Their conclusion, quoted at the end of Chapter I (Introduction) was that there is a continuing need for techniques like MAUT, but they need to be simple and easy to use.

Curry, Menasco, and Van-Ark (1991) used a multiattribute model to study how married couples make choices between two multiattribute dyads. They presented subjects with investment packages which had multiple attributes with two main categories of 'safety' and 'yield'. Kohli and Mahajan (1991) developed and tested a model for predicting pricing decisions. They used MBA students to rate the value/importance of apartment attributes provided in a survey developed by the authors. Troutt (1991) developed a theoretical model of how people make decisions on how much to add to their personal inventory when buying goods. As in all of the above papers, subjects were presented with a questionnaire and asked to rate various attributes in various categories.

Thomas, McDaniel, and Dooris (1989) employed a nominal group technique to develop attributes, and a MAUT technique for weighting attributes and assigning location measures in a study involving how strategic alternatives are selected in organizations. Finally, Shoemaker and Waid (1988) conducted a theoretical analysis of the probability of selecting one hypothetical college applicant over another given certain know information from a multi-attribute college admissions profile.

The common element in the previous papers was that they all focused on problems involving how people make choices and preferences given various bits of information, with varying importance weights, in consumer choice or decision contexts. It is clear, therefore, that MAUT is applicable and useful in a wide range of concrete situations which

include but go beyond program and training evaluation.

However, what the literature lacks is information about the administration of MAUT, the conditions under which it is effective or ineffective, the use and availability of stakeholders, the role of the evaluator, etc., for use in general by researchers and practitioners of MAUT, and for use in education or training evaluation in particular. In none of the above literature were any issues of relevance to the practitioners of MAUT. There is nothing about how to train a MAUT evaluator, nothing on how to select stakeholders, nothing on how to gain stakeholder commitment, nothing on how to conduct stakeholder meetings, nothing on how to present MAUT findings to users, and nothing on how MAUT reports should be written. These issues, and several more constitute what is referred to here as the MAUT process. It is this process that has not been examined up to now and is the object of study of the present effort.

We will turn now to a review of the application of MAUT to educational or training settings. Here too the focus of the research has been primarily on the technical aspects of MAUT and the results of its application. We will note again that, with the exception of the first two studies mentioned below, little of the MAUT process is mentioned in the application of MAUT to these settings as well, and the focus of all the studies is on the technical aspects of MAUT or on the program being evaluated. No studies explicitly address or aim at discovering more about the MAUT process itself.

Miller (1985) used MAUT to evaluate a talented and gifted student program in three school districts in Kentucky. A elaborate array of attributes and possible data sources was developed. Potential attributes were pulled from the literature and from individual interviews with a variety of stakeholders. The evaluator did report the



use of money and refreshments to get stakeholders to attend an attribute weighting session, however he still had difficulty getting desired attendance. He also described the use of a pizza buying example for introducing stakeholders to the technical aspects of MAUT. However, no concrete description of the example was provided for use by other practitioners. Miller also reported that the principals and superintendents who should have been involved were invited but could not free up their time. Other problems noted were difficulty in getting prior location measures by survey, and the extensive time and difficulty involved in developing measuring instruments to provide attribute measurements. Again, these problems of process were mentioned in passing and not the focus of the study.

Novak (1985) reported the use of MAUT to help set evaluation project priorities in a public school administration office. Staff members who reported to the Director of Evaluation for the school system provided desired attributes and were able, with some difficulty, to prioritize 55 projects. Novak notes difficulty in getting the real customers for the projects (e.g. the superintendent, cabinet members, etc.) to participate, in getting stakeholders to reach consensus on weights, and that the data quickly became very complex to manage. He recommended that a priority setting project like this requires a stable management environment which was lacking in this case. Again, although some issues of process were mentioned in the two previous studies, the evaluators set out to apply MAUT to concrete planning or evaluation problems rather than to examine process issues explicitly and directly.

Curlette (1991) reported the use of MAUT to aid school officials in the selection of instructional software. A committee of principals, teachers and instructional coordinators reached consensus on criteria

(i.e. attributes) to evaluate three different types of integrated software. Curlette reported that eight attributes were developed and that the programs were in fact able to be rank ordered on overall utility. Curlette also noted that commitment waned toward the end of the project and the MAUT decision strategy was not used by administrators. However, there was little description of the MAUT process used to select and train stakeholders, and no discussion of procedures that could be used to enhance stakeholder commitment.

Park (1979) applied MAUT to a strategic planning decision problem in the Education Department at the University of Pittsburgh. The objective was to develop a decision theoretic evaluation method for strategic planning.

The problem involved deciding which of several strategic policy directions to pursue for the coming year in the international student program of the University of Pittsburgh School of Education (International and Development Education--IDEP).

Six criteria, or MAUT attributes, were developed which represented the strategic goals of the Department. The entities to be evaluated were ten potential strategic directions statements aimed at achieving the six goals.

Three methods of arriving at utilities for the ten strategic directions were used and compared. The first was a simple rank ordering of the goals on the basis of how well they were believed to satisfy the six criteria. The second method was called the "plain utility" MAUT method, following Edwards et al. (1975). The third method was the decision theoretic, or expected probability method.

Park sought to determine whether (1) the three methods could, in fact, be used to solve the decision problem, (2) whether the three methods resulted in different decisions and which methods were more

similar in decision outcome than others, and (3) whether two different stakeholder groups (i.e., faculty and students) were similar in their perceptions of the importance of the various strategic alternatives.

A series of questionnaires and follow-up questionnaires were used to obtain attribute rankings, weightings, and location measures from two groups to stakeholders, students, and faculty.

The results indicated that (1) the three evaluation methods produced different results in terms of the final rank order of the strategies based on utility values; (2) the simple rank order method was more similar to the MAUT method than to the decision theoretic method, and the decision theoretic method was more similar to the MAUT method than to the simple rank order method, (3) the decision theoretic "expected utilities" method produced more variability among the final utility values than did the MAUT method; and, (4) there was a relatively high rank order correlation between final faculty and student ranks of the entities, suggesting high agreement between faculty and students in weighting the relative importance of the entities.

Vernon (1985) compared the MAUT procedure and a quasi-experimental impact evaluation method in a training evaluation setting. A Communications and Problem Solving course for Managers and Supervisors was offered in two different companies. Company E received the training (29 subjects total over two classes). Company C did not receive the training and was the control group (9 managers). The respective General Managers of each company were the two stakeholders in the MAUT portion of the study.

In the quasi-experimental study, a non-equivalent control group design was used. For the MAUT study, the two General Managers from the companies met independently with the researcher to collect attributes and location measures. The objective was to compare the conclusions

from the MAUT and quasi-experimental methods for similarities and differences.

In the quasi-experimental study, the dependent variable was "self" and "other" ratings on the ten training objectives. Each subject had an overall score which was computed by averaging all 40 ratings (i.e., self and three others' ratings on 10 objectives).

In the MAUT study, each manager generated his own evaluation criteria (i.e., MAUT attributes). The decision to be made in the MAUT design was whether the training should be continued (in the case of Manager E), or whether to start the training in the first place (in the case of Manager C). Utilities were thus needed for two decision alternatives: Option 1-"Training" and Option 2-"No Training". Thus, in assigning location measures (step 5 of the MAUT process) Manager E focused on what was the impact of training; and, Manager C focused on what would be the potential impact of training in his company. Utilities were then computed for each option, for each manager.

The results indicated that the quasi-experimental results of the evaluation were negative, i.e. no statistically significant effect for training.

In the MAUT study, the total utility values obtained from options 1 and 2 the manager of the company with the training (i.e., Company E) were compared. The value for the "Training" condition was 66% greater than the value for the "No Training" condition. The interpretation was that the training had 66% greater utility to the company, in the view of this manager, than the condition without the training. Thus, the two evaluation approaches yielded dramatically different results and implications for the decision maker.

Hawthorne (1985) compared the relative advantages of three methods of evaluating a management education program. They were:

(1) Cost/Benefit analysis, (2) MAUT, and (3) Impact analysis in the tradition of classical experimental and quasi-experimental design. All three methods were applied to the evaluation of one management training program to study the advantages and disadvantages of each method.

Participants in a Project Planning and Management course comprised the experimental group. A comparison group consisted of those eligible for the training but who were waiting to take the course. The setting was the management education area of a fortune 500 company.

In the quasi-experimental design, data from students were collected from pre-post and follow-up surveys, plus interviews three months after training.

In the MAUT design, eight company managers were used as stakeholders, and the group interview method was used for the MAUT process.

Stakeholders gave two sets of location measure ratings MAUT. First, they retrospectively rated trainee Pre-course proficiency on each attribute. Then they rated trainee Post-course proficiency, resulting in a Pre- and a Post-course MAUT utility index.

The MAUT results indicated a large increase in student knowledge was obtained from the course, but a lesser increase in attitude and application to the job was found. An overall utility score for student attitude, knowledge, and job performance (the three branches of the MAUT value/decision tree) for both Pre- and Post-course ratings was computed. The result was that the Post-course total utility index was 15% higher than the Pre-course index. The interpretation was that the training had an overall relative utility of 15%.

The results of the impact study were determined by the mean criterion score for the trained employees. Three months after the course, trainees completed self-ratings by survey on a series of items

reflecting the course objectives. The combined mean of these ratings was 2.82 on a five-point scale ranging from 1 = Not at all, to 3 = Satisfactory, to 5 = Very Much. The conclusion was that the course was ineffective in producing desired levels of productivity since the mean of 2.82 was slightly below satisfactory.

Hawthorne concluded that MAUT was a useful method which could be used by stakeholders to be as comprehensive or limited as their interests dictate, and that a different evaluation result was obtained as compared with the impact analysis. Whereas the impact analysis found no course effect, the MAUT analysis found a 15% greater course utility, and the cost-benefit analysis found a net gain per trainee of \$300 dollars.

In summary, this review determined that training evaluation is a systematic activity designed to improve training programs and to provide decision makers with information. It is a process aimed at reducing uncertainty about programs rather than a truth seeking enterprise in the classical hypothesis-testing sense. It calls for pragmatic approaches and methods. Utility for decision making, feasibility, and practicality are seen by current writers as necessary ingredients for the methods used. The classical experimental method is not well matched to the philosophical basis and the pragmatic needs of training evaluation for program improvement and decision making. The reason is that the purpose of the analytical experimental model is to test the hypothesis that whatever differences are observed between trained and untrained groups were greater than what is expected through chance. It is argued here that all this falls outside the needs of decision makers to make reasonable judgements of the comparative (multi-attribute) utility of various strategies.

MAUT, on the other hand, is viewed as potentially better suited to meet the needs of training evaluation. Advocates of MAUT claim that its advantages for training evaluation are many. Primary among them are (1) the decision-based focus, (2) the inclusion and recognition of decision maker values, (3) the quantitative structure that allows for comparisons among alternatives, (4) the ability to handle changing values, (5) the ability to handle experimental data, (6) the ability to produce structured evaluation information quickly.

Finally, the research to date on the use of MAUT for training evaluation is scant. Recent literature on the use of MAUT is focused on a variety of applications in the fields of consumer psychology, operations research, cognitive processing models, and in studying weighting methods for attributes. The literature is dominated by papers describing attempts to use MAUT for predicting decisions in a variety of contexts, and for predicting consumer choice patterns and consumer attitudes.

What appears lacking in the MAUT literature is research that develops, evaluates and describes the use of MAUT in the education or training program evaluation context. Further, the studies reviewed that have included the use of MAUT in training evaluation contexts addressed mainly the results of MAUT as compared with the results of quasi-experimental evaluations of the same programs. No study could be found that looked purposely and explicitly at what happens in the various MAUT steps, what problems emerge, and what conditions produce effective or ineffective MAUT evaluations, although some studies did make mention of problems encountered in the MAUT process.

The literature has also been scant on providing practitioners of MAUT with a supporting contextual administrative framework and protocol for the role of the evaluator, how to lead stakeholders groups, how to

question stakeholders to get them thinking about desired their own values and about program attributes.

The present study therefore focuses on bringing MAUT together with the need for new and better methods of training evaluation, with an eye toward gaining insights into the effective use of MAUT in this context. Finally, the research was conducted to provide a basis for the possible development of a MAUT user's guide as a product of the investigation.



## CHAPTER III

### METHOD

#### Overview

The primary purpose of this research is to develop, apply, and evaluate a MAUT approach to the evaluation of an organizational training program in order to create an opportunity to observe in some detail the process by which MAUT operates in a concrete situation. The theoretical interest was in assessing the potential of MAUT as an alternative evaluation model in the training environment. The practical interest was in studying the unique characteristics of MAUT in an applied situation to determine its strengths and weaknesses, and to discover the conditions under which MAUT is effective or ineffective.

The research plan involved applying MAUT to the evaluation of five training courses in a medium sized municipal government employee training department. Both quantitative and qualitative methods were used. Stakeholders were identified, criteria for evaluation (i.e. desired course attributes) were developed, attribute indicators and data collection instruments were developed, empirical data were collected, and the process and results were assessed by the stakeholders and by the investigator. The following discussion first describes the setting within which the training program took place, and then describes the procedures used in this study to evaluate the training program.

Setting - The setting will be described under four categories: 1) Description of purpose of training and the curriculum; 2) Training volume; 3) Measures and indicators used by the Personnel Services Department; and 4) Need Assessment.

1) Purpose and Curriculum. The Personnel Services Department provides a city-wide training and development program. The population served by the curriculum is approximately 2,000 city employees, including Police

and Fire Department personnel who are also eligible for training through the city-wide program.

The population of 2,000 includes a mix of staff, supervisory, and management personnel across a broad spectrum of occupational categories.

Except for specialized computer skills training, the curriculum offered through Personnel Services is the primary source of organizational training for this population.

The curriculum consists of four main categories of programs: 1) Management and Supervisory, 2) Communication, 3) Career and Personal Development, and 4) Miscellaneous. According to the 1990 - 1991 Curriculum Guide published by Personnel Services, approximately fifty-five (55) separate programs were offered in FY 91. The breakdown of courses by curriculum category is: Management and Supervisory (14), Communication (17), Career and Personal Development (20), and Miscellaneous (4).

The published goals of the curriculum are to: 1) improve job skills, 2) increase job satisfaction, and 3) enhance career opportunities within City government. Thus, employees enrolling in programs within the curriculum are assumed to have one or more of these goals in mind as short or long term training outcomes.

The organization uses consultants and local university staff as instructors for many of the courses offered (approx. 90%). It also draws upon the talents of experienced City employees to instruct in their specific areas of specialization.

According to the official training brochure, employees are encouraged to develop their own training plan, review the listing of courses, and decide which ones will be of value to them during the coming year.

Approximately six weeks prior to each class offering, announcements and course descriptions are posted on departmental bulletin boards. Course descriptions, date(s), times, and location are provided.

Employees must submit a nomination form to Personnel Services signed by their immediate supervisor, department head, and departmental training coordinator. Employees accepted into the course are usually notified by Personnel two weeks prior to the start of the class. Enrollment is on a first-come, first-served basis.

Most classes are set up to accommodate up to twenty students. To achieve economies of scale, Personnel strives to include as many students as possible, up to the twenty limit, in each class. Most courses are three to six hours in length and a small number are offered in more than one session.

While many of the courses delivered by contractors are often developed prior to being selected by the city, some tailoring of the programs does take place to meet the specific needs of employees.

For evaluation purposes, all courses include an end-of-class student reaction survey. However, other options such as pre-post knowledge testing, follow-up questionnaires, or interviews to gauge use on-the-job or organizational impact are typically not used.

The end-of-course survey includes an overall course rating, a three-part instructor rating, and a few open-ended questions regarding the achievement of training objectives, usefulness of content, time allotted for content, suggested changes and recommendations, etc.

2) Training Volume. According to the city budget document for FY 91, an estimated 228 individual training sessions were held in FY 90. Two hundred fifty (250) sessions are planned for FY 91. In addition, an estimated 2,937 students attended training in FY 90. Thirty one hundred

(3,100) students are projected for FY 91. These numbers also include computer science training students as well as those taking courses from among the four categories described above.

3) Measures. According to the city budget document for FY 91, the objective of the city-wide training program is "to arrange for training of the work force in microcomputer skills, management skills, and general skills." The document goes on to describe key indicators for the training program.

They are:

- o Number of training sessions conducted
- o Number of employees attending training
- o Average rating\* by training participants on post-training class evaluation form

\* On a scale of 1 to 5, with 1 = "Poor" and  
5 = "Excellent".

4) Needs Assessment. In January, 1989, and again during the same period in 1990, the Personnel Services Department conducted a city-wide needs assessment in order to determine what courses to offer the following fiscal year. The 1989 survey asked for: 1) an overall rating of training received by the respondent through the City-wide program; 2) a listing of the three types of training most needed by the respondent at that point in time (e.g. list 3 types of training that you believe would benefit you most at the present time); 3) a usefulness rating of various prospective course titles; and, 4) various questions related to the administration of the City-wide training program.

The 1990 survey took a different approach. First, the survey was aimed only at supervisors and managers. Second, it attempted to determine needs through the eyes of supervisors by asking them to rate the frequency in which various job conditions existed among their

people. For example, one statement read "The type of work my staff does requires public contact," and another read "Members of my staff must conduct meetings," etc. Thus, the survey attempted to determine training needs not by asking the needs question directly, but by asking key personnel what types of activity their staffs do and what types of problems/conditions are present. A frequency rating given to a standard list of statements was used. The intent was to determine problems that should be addressed through training.

#### Procedure for this study

Phase I consisted of stakeholder selection and the pre-workshop interview with individual stakeholders. The purpose of the interview was to gain confidence, establish rapport, and obtain support for the study. In order to get a head start on the development of desired course attributes, and to facilitate an understanding of the mechanics of MAUT, a series of specially constructed focusing questions were prepared and presented to each stakeholder.

Phase II consisted of the MAUT workshops. The purpose of the workshops was to elicit stakeholder values in the form of desired course attributes, and to rank and weight those attributes for later use in calculating MAUT utility values for each of the five courses under review. In addition, stakeholders were asked to develop organizational values and weight them on importance so that a strategic value score for each course could be obtained for later use in estimating total course impact.

Phase III consisted of the development of indicators and survey instruments for the MAUT attributes developed in Phase II. The Survey instruments were designed to collect data for use as measures of the indicators and to derive location measures for the various MAUT

attributes.

Phase IV was the distribution of surveys and data collection. Five courses were under review, each requiring the collection of survey data from students and their immediate managers.

Phase V was the processing of survey data and the calculation of location measures, utility values, and overall course impact indices.

Phase VI was the presentation of the MAUT results to stakeholders and a sample of former students.

Phase VII was an additional task constructed by the present investigator to clarify several issues which emerged during the study. This additional task involved a presentation of an imaginary memo to the stakeholder groups. The memo was written as if the Training Specialist in the Personnel Department had conducted a MAUT study and was briefing the Personnel Director. Stakeholder reactions to the process, results, and usefulness of the MAUT method were obtained. The justification for the use of this additional task is presented in section B (Stakeholder Availability/Time Commitment) in Chapter IV (Results).

#### A. Phase I -- Stakeholder Selection and Pre-Workshop

##### Interview

1) Stakeholder Selection. Nine (9) stakeholders were selected to participate as stakeholders for the MAUT process. The number was later reduced to seven (7) because of two drop-outs. The stakeholders consisted of the following mix of personnel and background:

- 1) a Training Specialist from the Personnel Services Department with a Master's degree in urban planning, who also served as internal coordinator for the study. This person was a mid-manager level in the organization and had intimate knowledge of the training courses.

- 2) the Director of Personnel Services with a master's degree in Education, executive salary level.
- 3) the General Manager of the Employee Relations Division within Personnel with a master's degree in management, executive salary level.
- 4) the Director of the Department of Human Services with a master's degree, executive salary level.
- 5) a Personnel Assistant/Secretary with a high school diploma (who later withdrew from the study due to resigning from the organization).
- 6) the Assistant to the City Manager (who later withdrew from the study due to time pressures and political factors to be discussed in Chapter IV) with unknown academic background, and with knowledge of City Manager level priorities.
- 7) a Police Department sergeant (academic level unknown) with much experience as a student in city courses.
- 8) a supervisor of warehouse personnel (academic level unknown) with much experience as a student in city courses.
- 9) a supervisor of information systems (academic level unknown) with much experience as a student in city courses.

To summarize, this sample of stakeholders included three executive salary employees, one supervisory employee with intimate knowledge of the training courses, one supervisory employee with knowledge of city manager level priorities, and three supervisors who, as students, had taken numerous courses from the city training curriculum.

2) Pre-Workshop Interview. The initial interview was conducted with each stakeholder individually. This was done in order to make sure that each stakeholder was aware of the specific purposes and tasks of this

project. Each stakeholders's concerns about being evaluated, expectations about the evaluation procedure and about the mechanics of MAUT were carefully addressed. The stakeholders were assured that they were not to be personally evaluated, that their judgements and values were absolutely critical to the success of the project, and that their particular orientation toward the training program is a necessary component of the full understanding of the strength and weaknesses of the program. The efficacy of this procedure in generating confidence, establishing rapport, and gaining support for the study is described in section B (Group dynamics) in Chapter IV (Results).

The second part of the initial interview with each stakeholder involved a series of four hypothetical questions (developed by the present investigator) designed to introduce each one to the task of describing desired attributes of a training program. The efficacy of this orienting task is also described in section B (Group Dynamics) in Chapter IV (Results) and served to provide a head start for the full attribute development process which was carried out in Phase II in which the stakeholders met for the first time as a group.

The hypothetical questions used to orient the stakeholders to the task of attribute development were:

- 1) Suppose you were a supervisor and your employee was attending a training class next week and you said to him/her "So long, I hope you enjoy the course. I hope it's a good one!"
  - a) What would a 'Good One' mean to you from a supervisory perspective?
  - b) What would a 'Good One' mean to you from a student perspective?
- 2) If you were the Director of Personnel and you had an important, costly decision to make about approving or disapproving the



offering of a particular course next year, what kinds of information about that course would you like to have before making that decision?

- 3) Describe the characteristics of the best training course you have ever taken.
- 4) Describe the characteristics of the worst training course you have ever taken.

The responses to the four questions were examined for overlap and redundancy, edited for correct grammar, and a revised list was produced to serve as a list of potential standard course attributes in the Phase II workshops.

#### B. Phase II -- The MAUT Workshops

A shift in the plan was necessary for Phase II when it was determined that it was impossible to assemble all nine stakeholders for a two- to four-hour workshop due to the loaded business calendars of the participants. It was decided to split the group in two--one group consisting of the three highest ranking stakeholders (in terms of pay level) in the organization, and the other consisting of the remaining stakeholders. Thus, the Personnel Director, the Director of the Department of Human Services, and the General Manager of the Employee Relations Division comprised the Executive Stakeholder Group. The other group was called the Mid-Management Stakeholder Group. The implications of the difficulty in assembling stakeholders will be discussed in Chapter V.

#### Executive Stakeholder Workshop No. 1

Participants assembled for a two-hour period for the first phase of the workshop. After an overview of the MAUT steps and the purpose of the study, stakeholders were asked to develop a list of five to ten organizational values and ranks and weight these using the ratio

weighting method described by Edwards et al. (1982). This procedure is described in pages 52 & 53 and is the same procedure used for the assignment of weights throughout the MAUT process, whether what is being weighted is values or desired course attributes. In the task currently being described, the weighted values allowed for the derivation of a strategic value score for each course for later use in the overall course impact measure, as will be seen.

The next task involved obtaining ratings from these stakeholders on the extent to which each course related to each organizational value. To do this, a matrix was created with organizational values going down the left-hand side and courses along the top. The instructions were to reach consensus on the extent to which each course related to (or supported) each organizational value. Stakeholders were told to rely exclusively on their current knowledge of the course content, and to assume an optimally designed course, course delivery, and the appropriate audience. In other words, the task was to consider the best case scenario for each course leading to an estimate of the upper limit of the extent to which the course related to or supported organizational values. A zero to 100 scale was used to assign the ratings, with zero defined as "No relation to/support of the value", and 100 defined as "Maximum relation to/support of the value."

After consensus was reached on the extent to which each course was related to each organizational value, the value ratings for each course were multiplied by the importance weights for each value respectively, then summed across values within each course to produce a unique strategic value score for each course. Workshop No.1 was then terminated as time ran out of the scheduled two-hour block.

Executive Stakeholder Workshop No. 2

The next workshop activity involved the development of the decision value tree of desired standard desired course attributes and their importance weights. It will be noted that in the lexicon of MAUT, attributes are called "twigs", twigs are grouped into "Branches", and a listing of branches, twigs, and attribute weights is called a "decision tree" or "decision value tree". See Edwards and Newman (1982) for further description.

Since the first workshop took much longer than planned to obtain strategic values and strategic value scores for each course, a survey was prepared by the investigator after the first workshop and distributed to the Executive Panel before the next workshop. The purpose was to define the next steps so that the stakeholders would come better prepared for the next workshop. The survey, if completely filled out by the stakeholders, would result in attributes, ranks, and weights for each stakeholder (see Appendix A for the survey). Workshop No. 2 took place two weeks after Workshop No. 1.

After the attributes were developed, they were ranked and weighted by the stakeholders within two branches (Process and Outcome Branches) using the same ratio weighting method used in weighting the organizational values. That is, attributes were first ranked highest to lowest on importance within each of two branches (Process and Outcome). Then, the lowest attribute within each branch was assigned a weight of ten (10), the next lowest a weight of 10 or higher depending on how much more important the attribute was than the first attribute, etc.

Weights within each branch were subsequently summed by the investigator and each attribute weight was divided by that sum to produce standardized weights within each branch.

Next, the Process and Outcomes branches as a whole were weighted

by the stakeholders. That is, stakeholders discussed which branch was least important, assigned that branch a weight of 10, assigned the other branch a weight relative to the first. These weights were in turn standardized by the investigator by summing the weights and dividing each weight by that sum. The result was a standardized weight for each branch, and separate standardized weights for each attribute within each branch.

To create attribute weights which reflected the combined weights of the each branch, "total" attribute weights were obtained by the investigator by multiplying each within-branch standardized attribute weight by its respective branch weight to obtain a final column of 15 weights which summed to 1.00.

A final activity was to obtain "prior" location measures for each course on each attribute following Edwards (1982). To do this, members reached consensus on a scale of zero (zero possession of the attribute by the course) to 100 (maximum possession of the attribute by the course) for each attribute for each course. The purpose was to obtain members' gut level assessment of the extent to which each course was currently realizing the desired attributes. These would be used later in the analysis to measure the extent to which prior location measures given by stakeholders were related to post-location measures produced through student and supervisor surveys.

#### Mid-Management Stakeholder Workshop

The Mid-Management Stakeholder workshop was to be modeled after the Executive workshop as a totally separate stakeholder group. However, the experience with the Executive Group with regard to the large amount of time required to conduct necessary tasks argued strongly for shortened procedures. It is also noted that just prior to the Mid-

Management workshop, two members of the group dropped out. One dropped because she was resigning from the organization, and the other because, as an assistant to the City Manager, she was too busy and asked that her time be formally requested in writing for further work on the project. The Personnel Director, not wanting at this point to raise any further attention to the project at the City Manager level, decided it was best to proceed without her. The implications of this for the politics of evaluation using MAUT are discussed in Chapters IV and V.

To save time for the workshop, the members were given the survey which had been given to the Executive Group between Workshop Nos. 1 and 2. This contained the potential desired course attributes. They were also given a look at the Executive Panel attributes as an illustration of the product they were expected to produce during the workshop. As a final shortening strategy, it was decided that the organizational values developed by the Executive Group would be used in the Mid-Management Group in lieu of developing new organizational values. The Mid-Management Group would, however, determine on its own the extent to which the five courses were currently related to/supportive of those organizational values.

The first workshop task, therefore, was to rate each course on the extent to which it is currently related to/supportive of the organizational values. This enabled the calculation of independent strategic value scores for the Mid-Management Group.

### C. Phase III -- Development of Attribute Indicators and Surveys

After desired course attributes were developed for both stakeholder groups, attribute indicators were developed. For each attribute in both decision trees, one or more survey questions were developed by the investigator as a proposed indicator for the attribute.

The nature and number of attributes dictated the number of surveys and survey items needed. A total of five surveys were determined to be required to provide measures for the location measures of all the attributes in both decision trees: a) Pre-Training Survey/Supervisor (see Appendix D); b) Pre-Training Survey/Student (see Appendix E); c) Training Feedback Survey (see Appendix F); d) Post-Training Survey/Student (see Appendix G); and, e) Post-Training Survey/Supervisor (see Appendix H).

Proposed indicators for each attribute were given to a representative from each stakeholder group for comment, recommendation, and approval.

#### "Post" Location Measures

Post-course attribute location measures (or, post-location measures) are measures of the extent to which a course is realizing the desired attributes as derived from the five survey questionnaires. All post-location measures were calculated as follows in both trees, except for the "learning" attribute: Item means were computed for each survey item within each survey. For example, the mean for item #1 of survey #1 (i.e. Pre-course/Supervisor) was computed for course #1 using all of the supervisors' responses to the survey. Likewise for course #2, and so on, resulting in five means for item #1 for survey #1. Then each mean was divided by 7, the maximum possible score for each mean. This transformation placed each mean on a standard percentage scale. Then, all transformed means for all the items associated with each attribute indicator were averaged to obtain an overall average transformed score. Finally, the score was rounded up to the nearest hundredths, and multiplied by 100 to remove the decimal. The result was the post-location measure for "that" attribute, for "that" course.

The "learning" attribute post-location measure was calculated as follows: items #1 (knowledge/skill rating after the course) and #15 (knowledge/skill rating before the course) from the end-of-class feedback survey were used. The percentage of students who indicated at least a one point gain (arbitrarily proposed as a standard by the investigator and accepted by the stakeholders) in knowledge or skill from attending the course was determined by comparing individual subject responses to items #1 and #15. In addition, post-course/supervisors survey items #3 and #4 were used, and the percentage of supervisors who indicated that their employee showed at least a one point gain in knowledge/skill from the course was also determined in a likewise manner. The two resulting percentages obtained from items #1 and #15, and from items #3 and #4 respectively were then averaged, rounded up to the nearest hundredth, and multiplied by 100 to remove the decimal.

#### D. Phase IV -- Distribution of Surveys and Data Collection

Each of the five courses under review required the collection of survey data from students and their immediate managers. During the period September, 1991, to January, 1992, survey data was collected from two classes of Supervision, two classes of Spanish, and one class each of Writing, Reading, and Cultural Relations. These were the only class offerings of these courses during this time period.

Pre-course/supervisors surveys were distributed via inter-office mail one to two weeks prior to the start of a course. Pre-course/student surveys were administered the first day of class directly in the classroom. Training feedback surveys were administered in-class and the conclusion of the course. Student and Supervisory post-course surveys were distributed via inter-office mail two to three months after the end of the course.

Courses varied in length from eight hours (Supervision and Cultural Relations) to thirty-two hours (Spanish) spread out over four weeks.

E. Phase V -- Prior and Post-Utilities, and Total Impact Scores

Prior utilities were computed by multiplying the prior location measures by their respective Total Weights for each attribute for each course. Total Prior Utilities were computed by summing individual attribute utilities within each course.

Post-location measures were derived from the survey data as explained earlier. Post-utilities were computed by substituting post-location measures for prior location measures and applying the same multiplication and summing procedures used to produce individual attribute and Total Prior Utilities as noted above.

In addition to the calculation of prior and post-utilities, correlations between prior and post-location measures were computed. A high correlation between stakeholder subjective judgment and the more objective student/supervisor survey data would lend support to the notion that stakeholder generated, MAUT workshop location measures were predictive of the survey data. If this were the case, the survey data would be redundant, and the workshop data could be used as a substitute for the survey findings. The practical implications of this would be that the surveys would not be necessary as long as the opportunity for stakeholder workshop measures was available. This would cut down significantly on the time required to do a MAUT training evaluation. To examine this relationship, Pearson Product Moment correlations were computed between the pre- and post-location measures for each course in the Executive Group. Only the Executive Group data were used because the investigator was present as facilitator during the development of



the prior location measures and thus had more opportunity to answer questions and guide the activity. This resulted in greater confidence being placed in the Executive group data. The parallel data acquired from the Mid-Management group was assembled without the presence of the investigator, as this is discussed further in section B in Chapter IV (Results).

#### Total Course Impact

Total course impact was developed from a formula suggested by Brinkerhoff (1988) for determining course impact. The theory behind the formula is that for training to have real organizational impact, it must teach important content (i.e. is relevant to organizational mission/goals/values), it must be well designed and delivered (i.e. have utility in terms of sufficient amounts of desired program attributes), and it must be delivered to students with a real need--the more students who need and receive it, the greater the organizational impact.

The formula used for overall impact was:

$$\text{Impact} = \text{SV} \times \text{TU} \times \text{Need}$$

Where: SV = Strategic Value

(i.e. content importance)

TU = Total Utility

(i.e. Quality in terms of design, delivery, usage,  
etc.)

ND = % of employees with a direct job  
performance need

Overall course impact was computed for each course twice: once using prior location measure data, and once using post-location measure data.

F. Phase VI -- Presentation of MAUT results to stakeholders and students.

One of the research interests was, "To what extent do the results of a MAUT evaluation and the process involved appear credible and useful to stakeholders and decision makers?" To address this question, briefings and interviews were conducted for the two stakeholder groups and a sample of students who had taken one or more of the subject courses from September, 1991, through January, 1992.

For the students, the briefing consisted of presenting the MAUT post-location measures for each course on a white board, describing the process for arriving at the numbers, and capturing their immediate reactions to the data in terms of whether the data matched their expectations, whether there were any surprises, etc., based on their experience with the courses.

For the Executive and Mid-Management Stakeholder Groups, the primary procedure for feeding back results was to compose an imaginary memorandum going from the Training Specialist to the Personnel Director (see Appendix J). The memorandum contained the results of the Executive Stakeholder Group MAUT Tree. The amount of detail would approximate that found in an executive summary of a technical report. Following the memo, a series of questions were asked of each survey respondent in an attempt to assess MAUT credibility and usefulness. The entire survey instrument appears in Appendix J. A face to face briefing was tried with the stakeholder groups. However, stakeholder availability was a problem in scheduling and in getting all stakeholders to meet for all required activities. Thus, the imaginary memorandum was developed. The efficacy of using an imaginary memorandum to evaluate MAUT and to get reactions and tests understanding of the data is addressed in section B (Difficulty of required activities) in Chapter IV (Results).

### Student Interview

Seven (7) students were selected to be interviewed. Lists were made of all participants from the five courses. Those who participated in more than one course were identified and contacted by telephone. Of the twelve contacted, seven agreed to be interviewed.

Students were asked four questions: 1) Describe how the course did/did not work for you in the sense of your profiting from the experience; 2) What were your specific positive/negative reactions to the course? 3) Describe your degree of learning/non-learning; 4) Give examples, if applicable, of how you have used the training on-the-job.

The purpose of these questions was to assist in determining how the results of the MAUT quantitative evaluation compared with the spontaneous student expectations of course merit/worth. This would be used as a gross validation of the MAUT results.

Students were also asked to respond to the post-location measures from the Mid-Management tree. Each student was asked to examine the post-location measures for the course(s) they received and tell the investigator whether the results were: a) as expected; b) lower than expected; or, c) higher than expected. For each student, a notation was recorded for each attribute, for each course taken using those three categories. For example, if a student took the Writing course and he/she concurred with the post-location measure for the Learning attribute (derived from the survey data), and "O" was recorded indicating the attribute score was "Okay" in the student's view; an "H" was recorded if the student felt the location measure was "too high;" and, an "L" was recorded if the location measure was considered "too low" by the student.

### Rank Order Analysis

At the beginning of the student feedback interviews, subjects were asked to rank order the five courses on the basis of total perceived value to the organization. The question asked was, "Assume you were called by the Personnel Director and asked to rank order these courses in terms of overall impact to the organization. Assume also that the Director's purpose in asking was to draw on your knowledge and experience in order to help him make planning decisions for next year. How would you respond?" All seven stakeholder group members were contacted to participate in this exercise as well.

Each subject ranked all five courses per the above instructions. A total of 14 rankings (7 stakeholders and 7 students) per course were obtained. A total of 14 X 5 (courses), or 70 rankings altogether were collected.

To examine the relationship between stakeholder and student rankings of course impact and course rankings determined by MAUT results, the data from all five courses were combined (N=70). Pearson Product Moment correlations were computed between stakeholder/student rankings of the five courses (i.e. Variable #1), and each of the following variables: Variable #2 - course rank on the basis of Total Utility Score (Executive Group tree); Variable #3 - course rank on the basis of Total Impact Score (Executive Group tree); Variable #4 - course rank according to Total Utility Score (Mid-Management Group tree); and, Variable #5 - course rank according to Total Impact Score (Mid-Management Group tree).

### G. Phase VII -- Evaluation of MAUT Process, Results, and Utility Stakeholder Opinions on MAUT

A survey was developed to assess the strengths and weaknesses of

MAUT as a process and to assess the utility of MAUT results.

All seven participants in the two stakeholder groups received the survey (see Appendix J).

The survey included an imaginary memorandum from the Training Specialist to the Personnel Director. The purpose was to try to simulate a real world situation and to get answers to key questions.

Although written in memo form, the contents of the memo were intended to include the amount of material that would otherwise be placed in an executive summary of a technical report on the MAUT evaluation of the five courses. The memorandum was followed by several open-ended questions about the usefulness of the MAUT process.

#### Evaluation of MAUT Process

Following the imaginary memorandum and subsequent questions, the survey included another section in order to get a more quantitative assessment of the MAUT process. In this section, a MAUT procedure was used to capture stakeholder assessments of the process.

It was decided that a MAUT evaluation of the MAUT process would be appropriate. An attempt was made by group interview to have the Executive Stakeholder Group develop their own attributes by which to evaluate the process. The question put to them was: "Given that a training evaluation process/method is desired by this organization, what essential characteristics (i.e., attributes) should it have, what products should it produce, etc.?"

The stakeholders were unable, however, to produce a consensus list of attributes. The apparent reason was that they had difficulty thinking of their training evaluation needs and values in terms of desired MAUT attributes. The investigator tried to assist by suggesting that they consider some bi-polar evaluation dimensions such as

qualitative/quantitative, simple/complex, scientific/intuitive, oral/written, or experimental/non-experimental.

When this failed to produce an acceptable list from the stakeholders, the investigator chose four evaluation criteria from the literature to serve as attributes: Utility, Feasibility, Propriety, and Accuracy (Joint Committee Standards for Educational Evaluations, 1981).

The investigator developed definitions of the four attributes by combining examples and definitions found in the original presentation of the Joint Committee Standards (Joint Committee Standards for Educational Evaluations, 1981). Stakeholders were asked by survey to rank and weight the four attributes, and then to rate the MAUT process, which they had participated in, using a 0 to 100 scale for each attribute. Instructions and instruments are found in Appendix J, part B. The results of this evaluation of MAUT is found in section A (Using a MAUT Procedure to Evaluate MAUT) in Chapter IV (Results).

#### Concluding Statements about Method

Since the purpose of this study is to examine what has been called here the MAUT Process, it is critical that the MAUT Procedure used to reveal this process be described in detail. The Process is an expression of the implementation of MAUT and it is clearly necessary that the procedures which were implemented be laid out in full. Thus, the step by step actions which stakeholders were asked to perform have been described in this Methods section in great detail. It is their reaction to, and interaction with, these steps that constitute the raw data of this study.

#### Limitations

This investigation is limited by certain design considerations and

by certain factors related to setting. One limitation is the use of a single investigator. With a single investigator directing the MAUT evaluation, the ability to attribute results to the MAUT process itself is limited by the interaction of the investigator with the MAUT method. That is, some outcomes and observations may be more a result of the particular style, performance, and competence of the investigator and less a result of the MAUT process under investigation. Safeguards against this threat were controlled in part through the use of tape recorded notes, an investigator's diary, and feedback on notes and observations obtained from stakeholders.

Additional limitations include the use of non-random single case study design, non-random selection of courses to evaluate within the setting, and the non-random selection of the time period in which the study was conducted within the setting. The setting selection for the study was a local training department convenient to the investigator and convenient to site personnel officials in that the Director of Personnel Services had an interest in expanding the training evaluation capability of the department. This could be a possible source of bias in that other training departments may have been less willing to invite the use of the MAUT method. The courses selected for evaluation were selected because they offered the greatest variety of course type and topic being offered in a time frame convenient to the investigator and the training department. This could limit the ability to observe whether MAUT is inappropriate for certain types of courses. In addition, the training department was undergoing particular scrutiny by top management just prior to the beginning of the study. This could have affected the behavior and attitudes of the stakeholders. Finally, the study is limited in its capacity to generalize by the small number of stakeholders and the limited range of their demographic properties.

## CHAPTER IV

### RESULTS

#### Introduction

This chapter presents the results in two sections. Part I follows the outline of the procedure section (Chapter III, Method) and presents the products of the various research phases. Many of these products were used in the study as materials and stimuli to produce other products and observations about the MAUT process. It was decided that the reader would gain the maximum benefit and understanding of the characteristics and use of MAUT, and would better understand the conclusions drawn from the evidence, if the products of the MAUT steps and supporting procedures were presented in this Results chapter as the context for findings rather than being presented as procedures in Chapter III (Method). The reader will also recall that one goal of the study was to provide the basis for producing a MAUT user's guide for training evaluation. It was decided that these products were also important as descriptive illustrations of the nature of MAUT in this setting.

Part II follows the outline of the categories of topics serving as focus areas for the study, as outlined in Chapter I (Introduction). It includes largely qualitative evidence and observations about the efficacy of MAUT and the administrative procedures used, and also includes references to other literature and the implications for the effective use of MAUT.

#### Part I

##### A. Phase I -- Stakeholder Selection and Pre-Workshop Interview

Phase I consisted of stakeholder selection and the pre-workshop interview. The purpose of the pre-workshop interview was to gain



confidence, establish rapport, obtain support for the study, and to get a head start on the development of desired course attributes to be obtained in the subsequent MAUT workshop.

The list of potential standardized course attributes compiled from the interviews is presented in Table 1.

## B. Phase II -- The MAUT Workshops

### Executive Stakeholder Workshop No. 1

Table 2 presents organizational values developed by the Executive Stakeholders with importance weights listed. The weights were obtained by having each stakeholder rank order the values on importance, then assign a weight of ten (10) to the lowest ranking value, a weight to the next least important value such that twenty (20) would indicate twice as important as ten (10), thirty (30) would indicate three times as important as ten (10), and so on. Final weights were derived by averaging the weights assigned by the three stakeholders for each value.

The Executive Panel had little difficulty developing or reaching consensus on the organizational values. With the investigator acting as facilitator, the group proceeded much in the following manner: a stakeholder would suggest an organizational value, another stakeholder would second the motion, a third would concur and offer another value, it would be seconded by yet another stakeholder, and so on. When the list reached seven values and no more could be thought of, the group went back over the list and did some editing to reach a final product. The relative ease with which the Executive Group reached consensus was the first clue to the investigator that stakeholder competency in terms of communication skills, the ability to deal with and conceptualize the notion of organizational values, etc. was exceedingly important in MAUT. More will be said in Chapter V about stakeholder ability.

Table 1

**Potential Standardized Course Attributes  
Derived from Individual Stakeholder Interviews**

---

The program creates favorable reactions that are useful in motivating, retraining, and recruiting greatly needed employees.

The program enhances the value of participating personnel by increasing their capacity to perform in future roles and situations not related to their current jobs.

The program increases important job performance; training graduates perform important tasks with greatly enhanced productivity.

Learning is assured through proper delivery level and style.

Training leads to use in one's personal life as well as on-the-job.

Student participation and discussion is part of the method; adult learners are kept interested.

Instructor(s) encourages feedback and student interaction.

Students learn from the other students as well as from the instructor.

Training makes the job easier for the students.

Should move at comfortable pace.

Instructor should know material well, know audience well, and tailor material wherever possible.

There is a legitimate need for the course.

Instructor takes a personal approach and uses many examples from his/her own personal experience to illustrate points.

Has a (measurable) impact on the workforce.

Training goals and objectives are met; course is faithful to its stated purpose.

Course is well received by students; students enjoy the experience.

There is a lively, energetic atmosphere in class.

Instructor inspires and motivates students.

Course includes useful handouts, books, etc. which help the student after the course.

A course should make you think; you should have homework; you should be challenged.

**Table 1 (continued)**

Empowerment of employee to leave training with a sense of being a stronger person in the workplace.

Empowerment of having a skill applicable to whatever employee does (in and out of work).

High cost benefit--within ten days I would expect more from the employee.

Return to workplace with a text, manual, or other important job aid.

High benefit/cost ratio (delivered in one day as opposed to one week).

Content coincides with department head needs.

Employee returns to the job with a heightened enthusiasm for his/her job.

Broader perception of job fit to the organization.

---

Table 2

**Organizational Values Developed by Executive  
Groups with Importance Weights**

Organizational Value	Importance Weight
Provide quality service to the citizens; be responsive, maintain a commitment to excellence.	.22
Provide cost-effective service; maintain a balanced budget.	.14
Provide quality administrative support to elected and top city officials.	.19
Practice/promote high ethical standards in both internal and external relations.	.12
Support affirmative action; support EEO and professional development for all employees.	.10
Provide a broad range of support services to the public.	.05
Provide fiscally conscientious management of city programs and services.	.18

The group then reached consensus on the extent to which each course related to each organizational value (see Table 2 for values). Then, the value ratings for each course were multiplied by the importance weights for each value respectively, and summed within each course to produce a unique strategic value score for each course.

Table 3 presents the strategic value scores for each course. In effect, the score represents the extent to which a course has the potential to have impact on the organization in terms of the seven organizational values.

Workshop No.1 was then terminated as time ran out of the scheduled two hours.

#### Executive Stakeholder Workshop No. 2

Only one of the three executives had completed the interim survey designed to explain the attribute development and weighting process. A second executive had partially completed it but got confused, and the third did not complete any portion of the survey. This finding cast doubt on the feasibility of doing these steps of MAUT by survey.

The stakeholder who did complete the survey was more prepared to develop/discuss attributes. He, therefore, became a natural leader for Workshop No. 2 in suggesting possible attributes. With the investigator acting as facilitator, attributes were offered by the first member (who had completed the survey), modified or seconded by another member, concurred with by the third member, and so on, much as what was done with the values development process. Very little controversy or debate was encountered. One member even remarked that this group reached consensus very well. It will be remembered, however, that the potential attributes listed in Table 1 were available in the survey as well as in the workshop and were referred to openly by the group as being helpful

**Table 3****Executive Group Strategic Value Scores**

---

<b>Course</b>	<b>Strategic Value</b>
Writing	35.95
Reading	31.34
Supervision	42.31
Cultural Relations	38.39
Spanish	41.38

---

in developing a final list of attributes. Also, the skill and ability of this group to deal with the task was carried over from Workshop No. 1.

Table 4 presents the final branches, twig attributes, standardized branch weights, and total weights produced in Workshop No. 2.

Table 5 presents the prior location measures for the five courses.

#### Mid-Management Stakeholder Workshop

Table 6 presents the results of the strategic value exercise for the Mid-Management Group.

During the strategic value task, the investigator observed that the Mid-Management Group had a noticeably more difficult time in understanding the assignment and in reaching consensus. Two members had difficulty in grasping that to give a low strategic rating to a course for a particular value did not necessarily mean that this was a bad course. Rather, it meant that the course simply did not relate to or directly support a particular organizational value. It was also apparent to the investigator that these particular members had lower verbal skills than the others. This, together with the difficulty in grasping the strategic value concept, suggested that perhaps there was a cognitive/conceptual ability level required in some of these activities which could affect the successful application of MAUT, to the extent that verbal and conceptual abilities come together in the evaluation process.

The strategic value exercise took about one hour and thirty-five minutes of a planned two-hour block. It was decided, therefore, to brief the group on the attribute development exercise rather than begin that next step and surely run out of time. One member then suggested that with the aid of some written instructions, the group could convene

Table 4

## Decision Value Tree - Executive Stakeholder Group

Branch/Twig	Weight	Standardized Weight	Total Weight
Process (Weight = .25)			
Content aims at important needs	75	.21	.052
Learning is insured/knowledge or skill is gained	65	.18	.042
High benefit/cost ratio	60	.17	.042
Effective, competent, knowledgeable instructor	45	.13	.032
Well received/enjoyed by students	35	.10	.025
Useful text, handouts, etc.	30	.08	.020
Learning from other students possible	20	.06	.015
Teaching method matches material and audience needs	15	.04	.010
Lively, energetic atmosphere	10	.03	.010
Outcomes (Weight = .75)			
Increases current job performance	65	.31	.232
Provides useful job-aids/tools	45	.21	.157
Students empowered w/heightened self-esteem confidence, competence, etc.	35	.17	.127
Job tasks made easier	30	.14	.105
Contributes to professional growth	25	.12	.090
Practical application beyond the job	10	.05	.037



Table 5

## Executive Stakeholder Group Prior Location Measures

Attribute	Course/Location Measure				
	(1)	(2)	(3)	(4)	(5)
<b>Process (Weight = .25):</b>					
Content aims at important needs	85	65	75	80	85
Learning is insured/knowledge or skill is gained	55	60	60	45	55
High benefit/cost ratio	50	50	75	55	75
Effective, competent, knowledgeable instructor	70	70	50	65	75
Well received/enjoyed by students	80	70	75	70	65
Useful text, handouts, etc.	65	40	40	40	55
Learning from other students possible	30	30	70	70	20
Teaching method matches material and audience needs	80	75	40	75	75
Lively, energetic atmosphere	20	40	80	50	70
<b>Outcomes (Weight = .75):</b>					
Increases current job performance	75	70	50	40	65
Provides useful job-aids/tools	65	45	40	40	60
Students empowered w/heightened self-esteem confidence, competence, etc.	75	65	60	65	65
Job tasks made easier	65	65	40	30	35
Contributes to professional growth	50	60	70	50	45
Practical application beyond the job	40	40	25	45	50

- (1) Writing Course
- (2) Reading Course
- (3) Supervision Course
- (4) Cultural Relations Course
- (5) Survival Spanish

Table 6

## Mid-Management Group Strategic Value Scores

Course	Strategic Value
Writing	68.41
Reading	59.47
Supervision	61.40
Cultural Relations	53.40
Spanish	44.21

on their own and develop attributes and weights. The investigator agreed to this proposal and provided the group with the same written instructions given the Executive Panel in between their two workshop sessions.

Subsequently, the group was not able to follow those instructions and required the investigator's assistance by telephone on two occasions. This finding casts doubt on the ability of stakeholders to function independently during this phase of MAUT.

Table 7 presents the results of the Mid-Management Group attribute tree in the same format as Table 4.

Table 8 presents the prior location measures for the five courses modeled after Table 5 in the Executive Group.

It will be noted that without the direct guidance and in-person facilitator presence of the investigator, there is a high risk that the stakeholders will fail to properly edit and evaluate their attributes. For example, note the apparent redundancy among some of the attributes and the misplacement of other attributes under the wrong branch in the Mid-Management tree. In the Process Branch (see Table 7) one could argue that 'Useful text' is really an outcome attribute since one cannot fully assess text use until after the training, assuming that use refers to use back on-the-job. For another example, 'Goals and objectives met' and 'Positive student reactions,' put in the Outcomes Branch, are arguably process attributes since both could be assessed immediately at the conclusion of the training course.

### C. Phase III -- Development of Attribute Indicators and Surveys

For each attribute in both decision trees, one or more survey questions were developed by the investigator as a proposed indicator for the attribute. The final list of attributes and associated indicators

Table 7

## Decision Value Tree - Mid-Management Stakeholder Group

Branch/Twig	Weight	Standardized Weight	Total Weight
Process (Weight = .45)			
Learning is assured through good design	30	.26	.117
Content related to needs	25	.22	.099
Training is participative interaction encouraged	18	.16	.072
Instructor inspires, motivates, generated energetic atmosphere	18	.16	.072
Useful text, handouts, etc.	13	.11	.049
Instructor takes personal approach - illustrates points w/personal examples	10	.09	.040
Outcomes (Weight = .55)			
Training improves job performance	24	.22	.121
Goals & objectives met; faithful to stated purpose	23	.21	.115
Student confidence enhanced	22	.20	.110
Positive student reactions; motivated employees	20	.18	.099
Training used in one's personal life	12	.11	.060
Course well received by students	10	.09	.049

Table 8

## Mid-Management Stakeholder Group Prior Location Measures

Attribute	Course/Location Measure				
Process:	(1)	(2)	(3)	(4)	(5)
Learning is assured through good design	100	100	100	100	100
Content is related to needs	100	75	100	75	100
Training is participative interaction encouraged	20	15	80	75	80
Instructor inspires, motivates, generated energetic atmosphere	50	50	50	50	50
Useful text, handouts, etc.	90	50	10	25	75
Instructor takes personal approach- illustrates points w/personal examples.	10	15	90	100	5
Outcomes:					
Training improves job performance	25	20	25	18	35
Goals & objectives met; faithful to stated purpose	90	90	90	90	90
Student confidence enhanced	100	85	75	60	40
Positive student reactions; motivated employees	50	25	75	50	50
Training used in one's personal life	50	50	10	60	10
Course well received by students	100	80	50	60	90

- (1) Writing Course  
(2) Reading Course  
(3) Supervision Course  
(4) Cultural Relations Course  
(5) Survival Spanish

(i.e. survey questions) is presented in Appendices B and C.

Tables 9 and 10 summarize the attributes and associated indicators in terms of survey and survey item number.

D. Phase IV -- Distribution of Surveys and Data Collection

Five courses were under review, each requiring the collection of survey data from students and their immediate managers.

Table 11 presents the response rate by survey type for the five courses.

E. Phase V -- Prior and Post-Utilities, and Total Impact Scores

Tables 12 and 13 present prior and post-location measures, and prior and post-utilities for the Executive and Mid-Management decision trees.

Prior utilities were computed by multiplying the prior location measures (see Tables 5 and 8) by their respective Total Weights (see Tables 4 and 7) for each attribute for each course. Total Prior Utilities were computed by summing individual attribute utilities within each course.

Post-location measures were derived from the survey data as explained earlier. Post-utilities were computed by substituting post-location measures for prior location measures and applying the same multiplication and summing procedures used to produce prior utilities (individual and Total) as noted above.

Pearson Product Moment correlations were computed between the pre- and post-location measures for each course in the Executive Group. The results are shown in Table 14. As can be seen, only one correlation was statistically significant and the combined course correlation between pre- and post-location measures was significant, but only at the  $p=.10$

**Table 9**  
**Executive Stakeholder Group Indicators**

Attribute	Indicator	
	Survey *	Item #
Content aims at important needs	a	2
	b	3
	c	12
Learning is insured/knowledge or skill is gained	c	1
	c	15
	e	3
	e	4
High benefit/cost ratio	c	3
Effective, competent, knowledgeable instructor	c	4
	c	5
Well received/enjoyed by students	c	6
Useful text, handouts, etc.	c	7
Learning from other students possible	c	9
Teaching method matches material and audience needs	c	10
Lively, energetic atmosphere	c	11
Increases current job performance	d, e	1, 1
	d, e	4, 6
	d, e	5, 7
	d, e	6, 8
Provides useful aids/tools	c	7
Students empowered w/heightened self-esteem, confidence, competence, etc.	d	3
	e	5
Job tasks made easier	d	8
Contributes to professional growth	d	9
Practical application beyond the job	d	7

- \* a = Pre-Training Survey/Student  
 b = Pre-Training Survey/Supervisor  
 c = Training Feedback Survey  
 d = Post-Training Survey/Student  
 e = Post-Training Survey/Supervisor

**Table 10**  
**Mid-Management Stakeholder Group Indicators**

Attribute	Indicator	
	Survey *	Item #
Learning is assured through good design	c	1
	c	15
	e	3
	e	4
Content is related to needs	a	2
	b	3
	c	12
Training is participative interaction encouraged	c	9
Instructor inspires, motivates, generated energetic atmosphere	c	4
	c	5
	c	13
Useful text, handouts, etc.	c	7
Instructor takes personal approach-illustrates points w/personal examples	c	14
Training improves job performance	d, e	1, 1
	d, e	5, 7
	d, e	6, 8
Goals & objectives met; faithful to stated purpose		
Student confidence enhanced	d	2
	e	2
Positive student reactions; motivated employees	d	4
	e	6
Training used in one's personal life	d	7
Course well received by students	c	6

- \* a = Pre-Training Survey/Student  
 b = Pre-Training Survey/Supervisor  
 c = Training Feedback Survey  
 d = Post-Training Survey/Student  
 e = Post-Training Survey/Supervisor



Table 11

## Survey Response Rate by Course and Survey Type

Course	Pre-Training/ Student		Pre-Training/ Supervisory		Training Feedback		Post-Training/ Student		Post-Training/ Supervisor	
	Surveys Out	Surveys In (%)	Surveys Out	Surveys In (%)	Surveys Out	Surveys In (%)	Surveys Out	Surveys In (%)	Surveys Out	Surveys In (%)
Writing	17	14 (82)	17	4 (24)	14	14 (100)	14	11 (79)	14	9 (64)
Reading	20	18 (90)	20	12 (60)	15	15 (100)	15	7 (47)	15	12 (80)
Supervision	40	35 (88)	40	23 (58)	36	35 (97)	36	25 (69)	36	22 (61)
Cultural Relations	25	21 (84)	25	19 (76)	22	20 (91)	22	13 (59)	22	16 (73)
Spanish	40	34 (85)	40	35 (88)	32	23 (72)	32	14 (44)	32	17 (53)

Table 12  
Executive Group Location Measures and Utilities

Attribute	Writing				Reading				Supervision				Cultural Relations				Spanish			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
	Content aims at important needs	85	4.4	85	4.4	65	3.4	80	4.2	75	3.9	84	4.4	80	4.2	82	4.3	85	4.4	71
Learning is insured/knowledge or skill is gained	55	2.5	73	3.3	60	2.7	86	3.9	60	2.7	64	2.9	45	2.0	63	2.8	55	2.5	73	3.3
High benefit/cost ratio	50	2.1	92	3.8	50	2.1	88	3.7	75	3.1	88	3.7	55	2.3	81	3.4	75	3.1	76	3.2
Effective, competent, knowledgeable instructor	70	2.2	95	3.0	70	2.2	93	2.9	50	1.5	93	2.9	65	2.0	90	2.8	75	2.3	93	2.9
Well received/enjoyed by students	80	1.9	91	2.2	70	1.7	90	2.2	75	1.8	88	2.1	70	1.7	83	2.0	65	1.6	86	2.1
Useful text, handouts, etc.	65	1.3	89	1.8	40	0.8	94	1.9	40	0.8	86	1.8	40	0.8	74	1.5	55	1.1	75	1.5
Learning from other students possible	30	0.4	91	1.2	30	0.4	75	1.0	70	0.9	83	1.1	70	0.9	88	1.2	20	0.2	82	1.1
Teaching method matches material and audience needs	80	0.8	93	0.9	75	0.7	90	0.9	40	0.4	87	0.9	75	0.7	81	0.8	75	0.7	82	0.8
Lively, energetic atmosphere	20	0.1	93	0.6	40	0.2	89	0.6	80	0.5	86	0.6	50	0.3	84	0.5	70	0.4	87	0.6
Increases current job performance	75	17.4	58	13.4	70	16.2	53	12.3	50	11.6	50	11.6	40	9.2	48	11.1	65	15.0	44	10.2
Provides useful job-aids/tools	65	10.4	88	14.1	45	7.2	94	15.1	40	6.4	86	13.8	40	6.4	73	11.7	60	9.6	75	12.0
Students empowered w/heightened self-esteem, confidence, competence, etc.	75	9.3	71	8.8	65	8.1	70	8.7	60	7.5	61	7.6	65	8.1	55	6.8	65	8.1	52	6.5
Job tasks made easier	65	6.9	76	8.1	65	6.9	76	8.1	40	4.2	68	7.2	30	3.2	58	6.2	35	3.7	47	5.0
Contributes to professional growth	50	4.4	81	7.2	60	5.3	67	5.9	70	6.2	77	6.8	50	4.4	56	5.0	45	4.0	54	4.8
Practical application beyond the job	40	1.4	65	2.3	40	1.4	73	2.6	25	0.8	61	2.1	45	1.6	60	2.1	50	1.7	41	1.4
TOTAL =	66.1		75.9		59.9		74.5		53.0		70.1		48.4		62.8		59.2		59.6	

a = Prior location measure  
 b = Prior utility  
 c = Post location measure  
 d = Post utility

Table 13  
Mid-Management Group Location Measures and Utilities

Attribute	Writing				Reading				Supervision				Cultural Relations				Spanish			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
Learning is assured through good design	100	6.5	72	4.7	100	6.5	86	5.6	100	6.5	64	4.2	100	6.5	63	4.1	100	6.5	73	4.8
Content is related to needs	100	5.4	85	4.6	75	4.1	80	4.3	100	5.4	84	4.6	75	4.1	82	4.4	100	5.4	71	3.8
Training is participative interaction encouraged	20	0.7	91	3.5	15	0.5	75	2.9	80	3.1	83	3.2	75	2.9	88	3.4	80	3.1	82	3.2
Instructor inspires, motivates, generated energetic atmosphere	50	1.9	95	3.7	50	1.9	92	3.6	50	1.9	91	3.5	50	1.9	87	3.4	50	1.9	91	3.5
Useful text, handouts, etc.	90	2.5	89	2.5	50	1.4	94	2.6	10	0.2	86	2.4	25	0.7	74	2.1	75	2.1	75	2.1
Instructor takes personal approach-illustrates points w/personal examples	10	0.2	92	2.0	15	0.3	89	1.9	90	1.9	93	2.0	100	2.1	82	1.7	5	0.1	98	2.1
Training improves job performance	25	5.8	56	13.0	20	4.6	51	11.8	25	5.8	48	11.1	18	4.1	46	10.6	35	8.1	42	9.7
Goals & objectives met; faithful to stated purpose	90	14.4	80	12.8	90	14.4	75	12.0	90	14.4	68	10.9	90	14.4	67	10.7	90	14.4	56	9.0
Student confidence enhanced	100	12.5	71	8.8	85	10.6	70	8.7	75	9.3	62	7.7	60	7.5	55	6.8	40	5.0	52	6.5
Positive student reactions; motivated employees	50	5.3	66	7.0	25	2.6	60	6.4	75	8.0	56	6.0	50	5.3	53	5.6	50	5.3	49	5.2
Training used in one's personal life	50	4.4	65	5.8	50	4.4	73	6.5	10	0.8	61	5.4	60	5.3	60	5.3	10	0.8	41	3.6
Course well received by students	100	3.5	91	3.2	80	2.8	90	3.2	50	1.7	88	3.1	60	2.1	83	2.9	90	3.2	86	3.0
TOTAL =	63.7		72.1		54.7		70.0		59.8		64.5		57.5		61.7		56.4		57.0	

a = Prior location measures  
b = Prior utility  
c = Post location measures  
d = Post utility

**Table 14**  
**Executive Group Correlation of Pre and**  
**Post-Utility Scores by Course (N=15)**

---

<b>Writing</b>	<b>Reading</b>	<b>Supervision</b>	<b>Cultural Relations</b>	<b>Spanish</b>	<b>Combined Courses (N=75)</b>
-.098	-.149	.313	.595 (p=.02)	.307	.20 (p=.10)

---

level. This indicates that, for this study, stakeholder consensus of course utility was only moderately correlated with location measures obtained through an independent sample of students through surveys.

Table 15 presents the same information for Mid-Management. No correlations were significant.

In conclusion, support was weak for the idea that prior location measures provided by stakeholders were positively correlated with post-location measures collected from students and supervisors. The interpretation of this is that the stakeholders saw the course differently from the students and supervisors at the attribute level.

The practical implication of this is that evaluators should proceed with great caution when attempting to rely on decision level stakeholder judgment alone, without including program participant views and the views of their managers. At the attribute, it appears that decision level stakeholder judgment is different from participant level judgment of course strengths and weaknesses.

#### Total Course Impact

Tables 16 and 17 present total prior and post-impact scores for the Executive and Mid-Management Groups respectively. The results indicate that for the post-impact scores there is a rather broad range of impact across the five courses, much broader than the range obtained for the post-utility. This is due largely to the variability among the courses in the estimated percent of the work force who have a job-related need for the training. What this points to is that a course such as Supervision can have a high strategic value and a relatively high utility score, but have a lower overall impact score due to a lower number of people in the work force who have a direct need for it.

On the other hand, a course such as Cultural Relations can have a

Table 15

Mid-Management Group Correlation of  
Pre and Post-Utility Scores by Course (N=12)

Writing	Reading	Supervision	Cultural Relations	Spanish	Combined Courses (N=60)
.04	.31	.14	.34	.21	.18

Table 16

Executive Stakeholder Group Utility/Impact Summary

Course	Prior Utility	Post-Utility	Percent Need	Strategic Value	Prior Impact	Post-Impact
Writing	66.1	75.9	61	35.95	1451.6	1665.9
Reading	59.9	74.5	73	31.34	1371.2	1705.8
Supervision	53.0	70.1	42	42.31	943.6	1246.0
Cultural Relations	48.4	62.8	82	38.39	1538.0	1738.0
Spanish	59.2	59.6	52	41.38	1275.0	1284.0

Table 17

Mid-Management Stakeholder Group Utility/Impact Summary

Course	Prior Utility	Post-Utility	Percent Need	Strategic Value	Prior Impact	Post-Impact
Writing	63.7	72.1	61	68.4	2661.1	3010.8
Reading	54.7	70.0	73	59.4	2376.7	3042.2
Supervision	59.8	64.5	42	61.4	1542.0	1665.0
Cultural Relations	57.5	61.7	72	53.4	2212.0	2375.0
Spanish	56.4	57.0	52	44.2	1299.0	1311.0



lower strategic value score and a lower utility score, but have a relatively high overall impact score due to the high course need. The nature of the content of Cultural Relations simply makes it relevant to more city employees than the other courses, thereby increasing the likelihood of impact.

It should be noted, however, that the true impact of a course such as Basic Supervision may be underestimated through the impact formula used here. The reason is that it is not just the supervisors who are affected by this type of course, but also the people they supervise. Thus, the relatively low total impact for Supervision is somewhat of a statistical artifact. The implication of this for the impact formula used here is that course need, as a factor in the equation, may need to be modified or supplemented to allow for courses such as Supervision. "Secondary Impact" or "criticality" may need to be the additional term in the equation.

F. Phase VI -- Presentation of MAUT Results to Stakeholders and Former Students.

Student Interview Results

Appendix I presents the investigator's summary of the student responses to the four questions described in the procedure section (p. 5). Table 18 summarizes the written summaries from Appendix I and categorizes the results. The numbers in the cells represent the number of students who offered positive or negative comments within the various categories, by course. The results indicate that the Writing course drew the most consistently positive responses, and the Cultural Relations course drew the most consistently negative. Supervision, Reading, and Spanish fall somewhere in between in order of positive responses -- highest to lowest.

Table 18

Number of Students ( Maximum = 7) Responding Positively or Negatively to Interview Questions

QUESTION CATEGORIES	Writing		Reading		Supervision		Cultural Relations		Spanish	
	No. Pos.	No. Neg.	No. Pos.	No. Neg.	No. Pos.	No. Neg.	No. Pos.	No. Neg.	No. Pos.	No. Neg.
1. "The course 'worked,' I profited from the training."	4		1		3			2	1	1
2. Specific positive/negative reactions were cited.	4			1	2	1		2	1	1
3. "Learning took place for me."	4			1	1	2		2	1	1
4. Applications of the training to the job were cited.	4			1	2	1		2	1	1

Students were also asked to respond to the post-location measures from the Mid-Management tree. Each student was asked to examine the post-location measures for the course(s) they received and tell the investigator whether the results were: a) as expected; b) lower than expected; or, c) higher than expected. For each student, a notation was recorded for each attribute, for each course taken using these three categories. For example, if a student took the Writing course and he/she concurred with the post-location measure for the Learning attribute (derived from the survey data), an "O" was recorded indicating the attribute was "Okay" in the student's view; an "H" was recorded if the student felt the location measure was "too high;" and, an "L" was recorded if the location measure was considered "too low" by the student.

Table 19 summarizes the findings of the student reactions to the post-location measures of the Mid-Management tree by course. The total number and percentage of "O's," "H's," and "L's" were tallied by course and presented along with percentages. Two points can be made from the analysis of student reactions. First, in all courses, the overwhelming majority of student reactions to the post-location measures indicated concurrence. That is, a belief that the location measure was about what the student would have expected the averages from the surveys to be. Second, the process of soliciting student reactions to post-location measures was useful in identifying clerical errors which, when corrected, further established the credibility of the results. When certain reactions to certain attributes were being recorded as consistently "too high" or "too low," the data were investigated. In some cases, it was determined that there were, in fact, clerical errors/typos in the data being shown to the students. Necessary adjustments were made to the data whenever typos were discovered. The

Table 19

**Summary of Student Reactions to Post-Location  
Measures, Mid-Management Tree**

<b>Category</b>	<b>Writing</b>	<b>Reading</b>	<b>Supervision</b>	<b>Cultural Relations</b>	<b>Spanish</b>
# "O's" (%)	31 (72)	7 (58)	16 (55)	17 (71)	14 (58)
# "H's" (%)	1 (2)	1 (8)	2 (7)	6 (25)	4 (17)
# "L's" (%)	11 (26)	4 (33)	11 (38)	1 (4)	6 (25)
<b>TOTAL =</b>	<b>43</b>	<b>12</b>	<b>29</b>	<b>24</b>	<b>24</b>

result was that the percentage of responses indicating concurrence with the results increased further.

To give the reader an example of the mistakes discovered and corrected, in the Writing course the students were presented with 91, 66, and 65 as location measures for the last three attributes in the tree. The actual numbers however for those attributes, in order, were 66, 65, 91. The same transposition errors were made in some of the other courses, such as 88, 57, 61 for Reading versus 57, 61, 88, etc.

Table 20 presents revised calculations from those presented in Table 19. The revision strategy for changing the numbers was as follows: If a student rated an attribute "too low," and the correct attribute location measure was, in fact, higher than the incorrect one that the student was responding to, the student rating was changed to "Okay" indicating concurrence with the rating. Likewise, adjustments were made if a student rating was "too high" and the actual correct location measure was lower than the one shown.

As Table 20 indicates, all "O" category response percentages for all courses, except Reading, increased when adjustments were made due to the clerical error. The results add further credibility to the credibility of the MAUT survey results.

### Rank Order Analysis

To test the hypothesis that MAUT results obtained by the survey of students and their supervisors are positively correlated with the independent, intuitive judgment of stakeholder group participants and a sample of students, the data from all five courses were combined (N=70). Pearson Product Moment correlations (NOTE: Spearman Rho Rank Difference Correlations is equivalent to Pearson Product Moment with an N as large as 70. See Guilford, 1978.) were computed between stakeholder/student

Table 20

**Revised Student Reactions to Post-Location  
Measures, Mid-Management Tree**

<b>Category</b>	<b>Writing</b>	<b>Reading</b>	<b>Supervision</b>	<b>Cultural Relations</b>	<b>Spanish</b>
# "O's" (%)	32 (74)	7 (58)	19 (66)	18 (75)	18 (75)
# "H's" (%)	1 (2)	1 (8)	1 (3)	5 (21)	2 (8)
# "L's" (%)	10 (23)	4 (33)	9 (31)	1 (4)	4 (17)
<b>TOTAL =</b>	<b>43</b>	<b>12</b>	<b>29</b>	<b>24</b>	<b>24</b>

rankings of the five course (i.e. Variable #1), and each of the following variables: Variable #2 - course rank on the basis of Total Utility Score (Executive Group tree); Variable #3 - course rank on the basis of Total Impact Score (Executive Group tree); Variable #4 - course rank according to Total Utility Score (Mid-Management Group tree); and, Variable #5 - course rank according to Total Impact Score (Mid-Management Group tree). To illustrate, seventy pairs of ranks comprised the correlation between variable #1 and variable #2. The first 14 pairs consisted of data from course #1, the second 14 from course #2, etc. Within each course, the first variable was course rank provided by each of 14 subjects; the second variable was course rank determined by one of the four other sources specified above (i.e. variables #2 through #5).

Table 21 presents the results of the correlations between stakeholder/student course rankings and each of the above variables. As can be seen, a low but statistically significant correlation was found between stakeholder/student rankings and Total Utility Scores in both decision trees.

#### G. Phase VII -- Evaluation of MAUT Process, Results, and Utility Stakeholder Opinions on MAUT

Appendix K contains the responses of the four stakeholders who responded to the stakeholder survey described in the procedure. Table 22 presents a summary of the raw data responses in Appendix K. The common theme reflected in the results is that MAUT has the potential for incorporation into routine training evaluation for the organization, but a streamlining and simplification is necessary for both those who must manage it and for those who must describe the findings.

Table 21

Pearson Product Moment Correlations Between  
Student/Stakeholder Rankings of Course Utility  
and Total Utility and Total Impact Scores (N=70)

Student/ Stakeholder Rankings vs. Total Utility Score--Executive Tree	Student/ Stakeholder Rankings vs. Total Impact Score--Executive Tree	Student/ Stakeholder Rankings vs. Total Utility Score--Mid- Mgmt. Tree	Student/ Stakeholder Rankings vs. Total Impact Score--Mid- Mgmt. Tree
r = .26*	r = -.148	r = .26*	r = .085

\* Sig. at .05 level



Table 22

### Highlights of Stakeholder Survey Questionnaire Responses

---

1. Comments to author of the memo:

- o Low utility courses should be dropped/replaced
- o Emphasis on personal goals of students should be lessened
- o Good recommendations! Go for it!
- o Useful material for planning training...especially course strengths and weaknesses
- o Technical terms should be better defined

2a. Credibility of results:

- o Credibility of numbers clouded by complexity of calculations
- o Difficult to understand! However, complex does not necessarily mean not credible
- o I believe! But, have trouble understanding Total Impact Score
- o Credible...if one assumes a comfort level with calculations

2b. Credible in eyes of top management:

- o Would be of modest interest, but not enough to override their own intuition
- o Their use of the results to make decisions doesn't hinge on credibility per se (the point here is that personal biases of officials is not always changed by "facts")
- o Results would be accepted if better packaged; recommendations probably not!

2c. Suggestions for additions to the memo:

- o How to simplify and institutionalize the method
- o Add description of how courses were chosen
- o Reduce technical process description to one paragraph, with more detail in an attachment
- o Better introduction (i.e. a road map) and glossary of terms, examples of calculations, etc.

3a. Effort to involve stakeholders:

- o Good cross-section of the City employees; instruments and complexity was overkill!
- o Use of stakeholders appropriate!
- o Not overkill! Reliability demands looking at all sides of issues
- o Appropriate!

3b. Development of indices and scores:

- o Too complicated!
- o Too time consuming!

3c. Strengths and Weaknesses of process:

Table 22 (continued)

## Strengths -

- o group discussion/participation
- o Thoroughness; covers all sides of issues
- o Interesting, challenging

## Weaknesses -

- o Complexity
- o Information overload
- o Not enough time spent in meetings (to enable better understanding)

## 3d. Recommendation of the process for future:

- o Has applicability if made simpler
- o Yes and No! Is better than in the past, but may be too much!

## 4a. Likely results of memo in real world:

- o Wayne ordered to pursue recommendations
- o Discussion with Training Committee; possible sharing with City Manager

## 4b. Sharing with City Manager:

- o Only if future decisions re training were challenged by City Manager
- o I would recommend sharing with City Manager
- o Only the bottom line would likely be shared
- o Probably would be shared after simplification by Personnel Director

## 4c. Use to defend training:

- o Could be used to defend training if packaged simpler
- o People would lose patience fast if not simplified
- o Could be an important factor in defense of training

## Other:

- o You should study how student ratings change over time
  - o We need this for all courses if it were shortened
  - o We should have met with you in whole or half day blocks
  - o How do I set a standard for a course utility score comparable to a five point scale where I can easily set a rational standard of 4 on a scale of five for my courses?
-

### Evaluation of MAUT Process

In order to get a more quantitative evaluation of MAUT, a MAUT procedure was used to capture stakeholder assessments of the process.

To do a MAUT evaluation of the MAUT process, attributes and importance weights were necessary. An attempt was made by group interview to have the Executive Stakeholder Group develop their own desired attributes for a program evaluation. The question put to them was: "Given that a training evaluation process/method is desired by this organization, what essential characteristics should it have, what products should it produce, etc.?"

The stakeholders were unable, however, to produce a consensus list of attributes. The main reason was that they had difficulty thinking in terms of desired evaluation attributes. The investigator tried to assist by suggesting that they consider some bi-polar evaluation dimensions, such as qualitative/quantitative, simple/complex, scientific/intuitive, oral/written, experimental/non-experimental, etc.

When this failed to produce an acceptable list from the stakeholders, the investigator chose four evaluation criteria from the literature to serve as four attributes: Utility, Feasibility, Propriety, and Accuracy (Joint Committee Standards, 1981). The investigator developed definitions for each attribute.

Table 23 presents the results of the MAUT evaluation of the MAUT procedure used in this study. The most uniformly consistent evaluation was for the attribute "Propriety" (i.e., LM standard deviation of 5). As expected from the data in Table 22, "Feasibility" was rated lowest in utility (utility score = 10.29). Finally, the highest overall utility score was for the attribute "Accuracy" (utility score = 19.13). It was concluded that of all the attributes, stakeholders found MAUT to be accurate useful, properly and professionally administered, and feasible,

Table 23

MAUT Evaluation of MAUT Process in Evaluation  
of Five Courses (N=4)

Attribute	Weight	SD	Standard Weight	Avg. Location Measure	SD	Utility Score
Utility	22.5	10.3	.28	61.25	11.3	17.15
Feasibility	22.5	8.2	.28	36.25	18.4	10.29
Propriety	10.5	.86	.13	85	5.0	11.05
Accuracy	23.75	4.1	.30	63.75	9.6	19.13

TOTAL = 57.62

in that order.

It should be noted that the feasibility issue was discussed with the Personnel Director and the Personnel Training Specialist. It was learned that in their view they do not have the clerical support necessary to administer five surveys for each of their courses on a regular basis in order to continue to do MAUT evaluations as routine evaluation method. This was the reason it got low ratings for Feasibility. The implications of this for MAUT training evaluation practitioners is strong and will be addressed in Chapter V.

## Part II

The qualitative evidence collected allowed for the discussion of MAUT along several categories of topics. That is, both the positive and negative evidence clustered within certain categories related to the value or utility of MAUT as a technique. Thus, the presentation will be organized around the following areas: 1) Stakeholder Availability/Time Commitment; 2) Difficulty of Activities; 3) Use of Surveys; 4) Group Dynamics; 5) Organizational Politics; 6) Utilization of Results; 7) Accuracy/Credibility of Results; 8) Reaching Consensus; 9) Miscellaneous.

### 1) Stakeholder Availability/Time Commitment

**Evidence:** Stakeholder availability for interviews and MAUT workshops was limited and difficult to arrange. This finding was noted from the very beginning of the study when the investigator requested that eight to ten stakeholders be provided for a two to four hour MAUT workshop. The investigator was told that four hours would be virtually impossible for most, if not all, of the stakeholders; that two hours would probably

be the maximum; and, that at least a two week notice would be required for the stakeholders. At that point the investigator decided to establish two stakeholder groups and to conduct preliminary individual interviews.

The first Executive Stakeholder Group meeting was held approximately one month after the initial decision to do so. Two hours were planned for the meeting. One of the stakeholders arrived twenty minutes late and another announced he would be leaving after an hour and fifteen minutes to attend to a previous appointment.

The group convened again within two weeks for another planned two-hour session. However, once again the stakeholder who left early at the first session announced he would be leaving after one hour. His leaving early meant that he would not participate in the prior location measure exercise which was ultimately done without him.

Finally, a meeting called several months later to brief the group on the MAUT results and to record their evaluations of the MAUT process was shortened due to two of the three stakeholders having to leave after only one hour. This resulted in the decision to develop the final MAUT survey containing the imaginary memo and the follow-up questions (see Appendix J).

A similar pattern was found in the Mid-Management Stakeholder Group. At the first meeting, one stakeholder was thirty-five minutes late...having to drive across town to attend. At the second meeting (not attended by the investigator at the groups' request), the same stakeholder was not able to attend at all. Finally, a results-and-discussion meeting similar to that held with the Executive Group was attempted by the investigator, but had to be split into two sessions with two stakeholders per session in order to accommodate stakeholder business schedules.

**Implications:** Assembling stakeholders for necessary meetings is likely to be a problem in other applications of MAUT in organizational settings. One solution would appear to be a stronger up-front commitment from the participants before the study, given directly to the investigator. In the current study, the investigator obtained a commitment, but it was indirectly through the internal coordinator, who, as a mid-level manager, had only a limited amount of the necessary influence that would be needed to obtain greater commitment. Another solution would be to create a formal evaluation team with the investigator as the lead person. All the necessary activities, including data analysis, would be performed as a group by the team. More will be said of this in the recommendations section.

**Evidence:** Another aspect of the time dimension involves the time actually used by stakeholders, by activity. The breakdown of time was: one (1) hour for the Preliminary Interview; two (2) hours for Development of Organizational Values and Rating of Courses on Same (it is acknowledged, however, that this was an "add-on" step, not mandatory for every MAUT evaluation); one (1) to two (2) hours for Attribute Development and Prior Location Measures; one (1) to two (2) hours for Development of Indicators; and, three (3) hours for Feedback of Results.

The preliminary interview time prior to the workshops appeared to be a vital part in order to get stakeholders thinking about what, in their opinion, constitutes worthwhile training. It appeared that the more time that could be provided for this type of "orientation," the better because stakeholders are not necessarily used to thinking about programs in terms of attributes.

The organizational values session included an overview of MAUT, development of organizational values, ranking and weighting of values,

and ratings of the extent to which each course related to each of the organizational values.

There were two reasons why this step took all of two hours. The first was that none of the stakeholders had ever thought of their organization in terms of mission and values. Consequently, a rather large amount of time was spent giving examples of organizational values, how they are worded, etc. The second was that rating the extent to which each course related to each organizational value required a certain level of familiarity with the course content. One stakeholder had not taken many of the courses under review and thus had difficulty relating to the exercise. As a result, he had to defer to the other stakeholders on several occasions while trying to reach consensus and the resulting discussions were time consuming.

The Attribute Development and Prior Location Measures session included developing MAUT course attributes, ranking and weighting them, and rating the courses on the extent to which each course was thought to currently possess each attribute. The full two hours was required for this activity. The reason the full two hours was used was that one stakeholder did not read the written survey instructions that were sent out prior to the meeting. As a result, the group was slowed while this stakeholder got familiar with the list of potential attributes. Once this was done, the time and pace of the discussions and decision making appeared appropriate and free of problems.

The Mid-Management Group met independently (i.e. without the investigator) after they had been briefed earlier on procedure and were provided a handout of instructions (i.e. the same handout used by the Executive group between the organizational values exercise and the attribute development exercise. See Appendix A for the handout). The group had difficulty interpreting the written instructions and halted



activity until they could reach the investigator by telephone. They reconvened at a later date and received instructions by telephone again from the investigator. It, therefore, did not appear that this step could be done by survey or without direct guidance by the evaluator.

Proposed attribute indicators were developed by the investigator for some, but not all of the attributes in both decision trees. This process took about one hour. The proposed indicators were then given to a representative from each of the two stakeholder groups. They reviewed and edited the indicators and developed new ones as appropriate. Each worked on this activity independently. Each reported that they spent about one hour on this task. This task appeared easy, simple, and straightforward.

An oral briefing to feed results back to Executive Stakeholder members was conducted. Two members had to leave after one hour (this led to the decision to develop the imaginary memo approach to feed back results and get reactions. See Appendix J). The oral briefing was conducted and completed with at least one stakeholder. It included an overview of the procedure, and a walk-through of various handouts containing MAUT attributes, prior and post-location measures, impact scores, and attribute and total utility scores prepared on a Lotus spreadsheet. It also included time for responses to questions asked by the investigator designed to evaluate the MAUT process and results, assess the probability of use of the results, etc.

The briefing, questions and answers, and the structured evaluation questions took the full two hours of allotted time. Five sets of MAUT pre- and post-location measures and utility data (1 set per course) was an immense amount of information to explain. Only highlights were covered. Several questions were asked about the various tables of data. The MAUT decision trees, attributes, weights, location measures, etc.

and complex when arrayed in tables. It requires slow explanation and generates questions, which add time. In retrospect, three hours is recommended because the briefing was very rushed, partly for the benefit of those who had to leave early. In addition, direct feedback obtained during the briefing indicated that stakeholders were overwhelmed by the data when it was presented in one large dose in a briefing. More time would allow for a slower delivery and for greater retention by the stakeholders.

**Implications:** The implications of this set of findings are simple and direct. MAUT is an orderly sequential process which takes one (1) to three (3) hours per step, as determined by the present study.

**Evidence:** A separate source of time related evidence was found among the negative responses received from stakeholders in the stakeholder survey at the conclusion of the study. In short, stakeholders complained of information overload and demands on stakeholders time. Many of the "overload" responses were directed at the amount of statistical information included in the imaginary memo/report in the stakeholder survey. Comments about time were aimed at both stakeholder time (i.e. the workshops), interviews, and questionnaires, and also the overall length of time over which the study was conducted.

**Implications:** The significance of these findings is great. The value of MAUT as an evaluation method is directly related to the credibility of the results it generates, and the ease with which decision makers can understand and explain the results. If MAUT cannot be easily understood by the stakeholders and other decision makers, MAUT evaluations will fail to get used by evaluators, or by decision makers and program

directors.

In addition, these findings also suggest that greater effort needs to be made to clarify terms and other technical aspects of MAUT than was done in this study. Finally, less information is advised for written reports of MAUT results.

## 2) Difficulty of Activities

**Evidence:** One of the tasks presented to the stakeholders at the beginning of the workshops was to imagine that they were the Personnel Director in charge of Training and that they had to make an important decision about the future of the courses under review. In that scenario, the question was: "How comfortable are you currently, on a 100 point scale (100 = maximum comfort, 0 = minimum comfort), with the amount and level of information you have about the merit/worth of the various courses?" Another question was "What evaluation-type information would you like to have about the courses to feel good about making your decision?" The first question was asked to assess baseline uncertainty level against which to compare uncertainty levels subsequent to the MAUT evaluation in order to assess the utility of MAUT for reducing uncertainty (and therefore aiding decision making). The second question was asked to get the group thinking about decision criteria that could be used eventually for MAUT attributes.

The Executive Group was able to relate conceptually to the hypothetical scenario of the question.

The Mid-Management Group, on the other hand, had difficulty relating to the questions. For example, one stakeholder repeatedly tried to answer the first question (i.e. re uncertainty level) by talking about the merit/worth of the courses. The investigator concluded, through this and by other comments and questions, that this

stakeholder simply could not understand the question and would not in the time available.

As another example, the reader will recall that one task of the Mid-Management group workshop was to rate each of the five courses on the extent to which the course content related to each of the organizational values developed by the Executive group. The task was to reach consensus for each course, for each organizational value, by assigning a rating from 0 to 100 (100 denoted maximum relation of course content to organizational value). As in the earlier example, the Mid-Management Stakeholders had problems relating to the task. Two stakeholders in the Mid-Management Group repeatedly showed by their questions and comments that they viewed the rating as a measure of course merit/worth rather than as a measure of the extent to which course goals were congruent with the various organizational values. The investigator needed to remind the group several times of the purpose of the exercise.

**Implications:** The implication of these experiences is that task difficulty in MAUT can be an important factor, and that stakeholder skill/ability to deal with the tasks is important for planning a MAUT evaluation and selecting stakeholders to participate. Moreover, educational level is likely to be an important factor in the stakeholder effectiveness due to the strong cognitive component of the tasks. The fact that more difficulty was experienced by the Mid-Management Group than by the Executive Group, and given the higher overall ability, educational level, and salary level/status in the Executive Group, it is plausible that educational level is an important factor in stakeholder effectiveness.

**Evidence:** Another source of evidence for discerning the difficulty of MAUT was the final stakeholder survey. A prominent theme from the survey results was that the credibility of the MAUT results was clouded by the complexity of the calculations presented in the imaginary memo. One stakeholder said that the complexity did not detract from the credibility per se, but it detracted from his ability to sell the information upward in the organization.

The efficacy of feeding MAUT results back to stakeholders in the form of an imaginary memorandum, however, appeared to have great usefulness as a process to gauge the extent to which stakeholders are truly understanding what is being communicated. It is noteworthy that the feedback that the evaluator was receiving during the oral presentation of MAUT results (which was cut short due to departing stakeholders) did not indicate a lack of understanding of the results. However, responses to the memo containing highlights of the results drew comments about complexity of the information. It is plausible to conclude that the feedback from stakeholders in the face to face setting was unreliable due to a reluctance of stakeholders to confess confusion in a group setting. The implications of this for the positive use of written results by evaluators as a tool to both communicate and get feedback on understanding are straightforward.

**Implications:** The significance of these findings is great. As is the case with time issues and MAUT, the value of MAUT as an evaluation method is directly related to the ease with which decision makers can understand and explain the results. If MAUT cannot be easily understood by the stakeholders and other decision makers, it will probably fail to get used by evaluators, decision makers, and program directors.

These findings suggest that greater effort needs to be made to clarify terms and other technical aspects of MAUT than was done in the present study. In addition, less information is advised for written reports of MAUT results.

### 3) Use of Surveys

**Evidence:** Surveys were used to develop attributes when time ran out in the initial workshops for the executives and mid-managers. Although the evidence indicated that the survey (see Appendix A) was not overly complex, only one executive worked through the survey in preparation for the next workshop. The mid-managers telephoned twice for explanation of specific points. It was clear that reliance on survey methods for this task is tenuous, if not appropriate. Further, the quality of the products generated by the survey left much to be desired.

**Implications:** This evidence suggests that for the primary core activities of MAUT -- i.e. attribute development and weighting -- the use of surveys is not advisable. Although the step by step requirements appear simple to the evaluator, they are of sufficient complexity to discourage and/or pose difficulty for stakeholders.

### 4) Group Dynamics

**Evidence:** In a special survey of stakeholders at the end of the study, it was reported that positive group discussion/participation was one of the primary strengths of MAUT.

In addition, one on one interviews which preceded stakeholder group exercises were quite useful for establishing rapport, introducing the study and MAUT, and in eliciting initial ideas about the expected value of training. These ideas were subsequently used successfully to

help develop group consensus on desired course attributes.

However, in the group activities that followed the interviews, some stakeholders behaved much differently than in the individual interviews. In the group scenario, where the tasks were to develop, rank and weight desired course attributes, there were noticeable differences in how stakeholders related to the activities and to one another. In the executive group, boss and subordinate relationships were clear inhibitors to brainstorming. The Personnel Director was deferred to frequently by his subordinate, the Employee Relations Manager, who was clearly more outspoken in the individual interviews. The Personnel Director, very sure of himself and decisive in the individual interview, was clearly more tentative and less of a leader in the presence of the Director of Human Services. To an outside observer, it appeared that the Director of Human Services had more status in the organization. Yet, in actuality, the Director of Human Services reported directly to the City Manager just as the Director of Personnel.

It was later learned that the Director of Human Services had at one time worked at the University of Michigan doing program evaluation work.

**Implications:** A plausible conclusion from this is that formal organizational status, education, and work experience all were factors which effected the dynamics of the small group MAUT stakeholder activities, and which effected the success and quality of the process. The practical implications are that evaluators should be sensitive to the potential barriers that can hinder the success of MAUT group activities in organizations. The more the evaluator can learn about the organization through talking with or interviewing a variety of people

prior to forming stakeholder groups, the better he/she is able to form productive groups of stakeholders.

#### 5) Organizational Politics

**Evidence:** Political factors came into play in this study. The evaluation was constrained by the lack of involvement of the City Manager and her immediate staff who, it turned out, was the real decision maker for training. But it was the Personnel Director and his staff assistant who had an interest in the evaluation. The Personnel Director also had an interest in controlling the degree of involvement of the City Manager. Although the Personnel Director was openly interested in exploring the quality of training, he nevertheless controlled the study by not allowing the City Manager to get involved...allegedly because her controlling personality would cause her to control and micro-manage the process.

Political considerations have also dictated the type of training courses being offered in the past. For example, the City Manager, in attempting to be responsive to a community activist group called the Citizens for Responsible Taxation (CRT), recently cut many courses due to the belief that the courses were (or would be) perceived by the CRT as "fluff", i.e., a waste of employee time and taxpayer money. Most courses falling into that category did not (by title) appear to be directly related to any particular job, or did not appear to contribute directly to the improved job performance of any group of workers.

**Implications:** It is always important for evaluators to be aware of the political factors that will effect or influence their evaluation. This is especially important in MAUT where the ideal is to include all relevant stakeholders in the process. Evaluators need to realize that



their study will be as effective as it is able to include relevant stakeholder values and/or attributes. If the real stakeholders cannot be formally included, alternative ways of getting their input need to be explored.

#### 6) Utilization of Results

**Evidence:** Just as those who write about the utilization of program evaluation recommend involving the real decision makers as stakeholders in the design of an evaluation study (Patton, 1980), so too the advocates of MAUT recommend that those who make program decisions be involved in at least the attribute development process. Those persons, and the decisions which they have to make, are the target beneficiaries of the MAUT process. Therefore, the question of who/whom makes the real decisions about training, and, whether they can participate as stakeholders, is important.

In the present study it was learned that the "real decision maker" for training (at least recently) was the City Manager. The City Manager, in conjunction with a special Training Committee, held power and influence over specific courses offered by the City. However, the City Manager did not participate in the study and only two members of the Training Committee participated as stakeholders (i.e. two of the members of the Executive Stakeholder Group, as City Department Heads, were members of the Training Committee). Involvement of more Training Committee members or the City Manager was not feasible in practice because, again, of the severe time constraints posed by their executive business calendars. In the case of the City Manager there was the added constraint discussed earlier, i.e., that the Personnel Director felt that the inclusion of the City Manager would unduly bias or control the study.

The actual decision process for determining what optional/elective courses are offered by the City, it was learned, is [typically] as follows: 1) The Personnel Director, or associated staff, will identify a potential course need through some formal or informal needs analysis, 2) a proposed course description and design will be prepared, 3) the proposal will be submitted, through official channels, to the City Manager for approval, 4) the City Manager weighs his/her own intuitive beliefs about the value/need for the training against any relevant political and budgetary considerations, and approves/disapproves the offering of the new and/or continued offering of a previously offered course. Each course and each offering (i.e. specific class offering) is approved at the City Manager level.

**Implications:** The implications of how training courses are offered and who the real decision makers are can place constraints on the extent to which MAUT evaluation results will get used in an organization. Utilization of results was constrained in this study by the lack of participation of the City Manager.

**Evidence:** Another line of evidence on utilization of results came from the final stakeholder survey. Stakeholders indicated that there would be a slim chance that the MAUT results would be seriously included the decision making about the courses, in part because the City Manager and other top officials base many training decisions on their intuitive beliefs about the value of the courses.

It was also noted by some stakeholders that the MAUT results, although credible, would not override any inherent bias (held by the City Manager) for or against the training courses. The incorporation of organizational values into the evaluation equation was expected (by the

investigator) to draw particular interest and appeal from top management. The findings from the stakeholders, however, suggested that the MAUT design and results would not be sufficient to override the intuitive decision making style and practice of top management.

**Implications:** The implications of this are that MAUT evaluations of training need to involve the primary decision makers if the results are to influence decisions.

The environment for the present study did not allow for the inclusion of the primary decision maker. The consequence of this is that the MAUT results will probably get little use in decision making above the level of the Director of Personnel. The problem however is that in this organization the Director of Personnel is only one among many participating in decisions about training.

**Evidence:** One reason cited by the stakeholders for not wanting to share the results with the City Manager was that the report (i.e. the imaginary memo) was too complex and required major simplification to be understood.

**Implications:** The significance of this finding is, of course, very great and speaks directly to the value of MAUT and its utility as an evaluation method. In retrospect, much could have been done to provide definitions of terms and other aids to the reader to make the imaginary memo easier to understand. For evaluators, the implication is that extreme care should be taken to simplify written documentation of MAUT results.

### 7) Accuracy/Credibility of Results

**Evidence:** In the special survey of stakeholders at the end of the study, stakeholders reported that credibility was among the many positive characteristics and strengths of MAUT. In addition, stakeholders rated the accuracy of the results the highest of the four evaluation criteria. Accuracy as a dimension of the quality of MAUT as an evaluation method received the highest utility score (19.13). Utility, Feasibility and Propriety were the other three dimensions (see Table 23). This indicates that the credibility of the results was perceived relatively high by the stakeholders.

**Implications:** The theoretical implication of this finding is that MAUT does have value as a training evaluation method in the eyes of stakeholders. That value is transported mainly by the involvement of relevant stakeholders, who provide relevant evaluation criteria and importance weights.

**Evidence:** Additional evidence comes from the sample of students who were shown the results of the post-location measures for the Mid-Management tree. Their reactions as to whether the location measures appeared plausible were recorded. The purpose was to obtain a face validity check on the data. The results indicated that the students concurred with the data in the majority of instances, and, where they didn't, clerical errors were usually found. When corrected, these errors further increased the number of instances where agreement or acceptance of the results was obtained (see Tables 19 and 20).

**Implications:** The theoretical implication of this finding is that MAUT produces results which have face validity with key stakeholders, such as

program participants.

The practical implications are that: a) MAUT results should be routinely fed back to sample of former students to establish credibility, b) doing so serves as an error check on a process which can be vulnerable to clerical error because of the large number of bits of information being reported, and, c) doing so is consistent with the notion of doing MAUT with a formal evaluation team in that a team discussion of the results would be standard procedure in that scenario. The evaluation team approach would, in turn, have other advantages, such as greater stakeholder commitment of time and attention. More will be said of this in the recommendations section.

#### 8) Reaching Consensus

**Evidence:** Consensus was more easily reached in the Executive Stakeholder Group than in the Mid-Management Group. The Executive Group stakeholders followed an orderly pattern whereby a suggestion would be made by one party; a second party would comment and either modify the suggestion or second the motion; the third party would usually second one or the other of the previous suggestions, and consensus would be reached on the item.

The Mid-Management Group had noticeably greater difficulty reaching the consensus. The main conflict appeared to come in the exercise where the members were asked to rate the extent to which each course related to each of the organizational values developed by the Executive Group. The consensus process was hampered by the difficulty members had understanding the instructions.

**Implications:** It seems plausible to conclude from these observations that the consensus process in MAUT activities are clearly influenced by

the ability of the stakeholders to conceptualize what is required and to verbalize their thoughts in order for the consensus process to work. In this particular case the Mid-Management Group was hampered by difficulty in understanding the task, but also showed signs of never reaching consensus when the task was understood. In the latter case, collecting individual ratings, computing the mean and analyzing/monitoring the variance would be recommended.

#### 9) Miscellaneous Issues

**Evidence:** Both the Executive and Mid-Management Stakeholder Groups, after being briefed on the mechanics of MAUT and definitions of the terms "Branch", "Twig", etc., decided that their course attributes would fall into two main categories--Process and Outcome. Process refers to those aspects of training which pertain to the delivery of training or the training "process" itself. Outcome refers to aspects of training which are manifested only after the training is conducted, such as application of the training on the job.

The systems view of training, and the concept of training as a cyclical process of analysis, design, development, delivery and evaluation is well documented (Brinkerhoff, 1987). The fact that both stakeholder groups, knowing little if anything about the systems view of training, reasoned that in-class process attributes should be kept separate from post-class outcome attributes makes it plausible to conclude that at least these two stages of the training cycle make natural divisions for attributes in a MAUT evaluation of training.

**Implications:** The implications of this is that in future MAUT training evaluations it may be advisable to explicitly direct stakeholders to develop attributes within those two categories on the theory that those

are natural and therefore consistent categories for training evaluation.

**Evidence:** Some evidence was found which suggested that executive level stakeholders view training differently than lower level stakeholders. There were noticeable differences between the two decision value trees of the Executive and Mid-Management groups, as well as some similarities. One difference was that the Executive Group placed a much heavier weight on the Outcomes branch of the tree. Another difference was that, in general, the Executives rated the courses lower during the prior location measure exercise.

In contrast, higher weights were given the Process attributes (see Table 13), and the higher prior location measures (see Table 13) were given by the Mid-Management.

As for similarities, the two groups were similar in the types of attributes they developed and in the types of branches they selected for their attributes. The significance of the similarities, however, is tempered somewhat because the Executive attributes and branches were indeed shown to the Mid-Management stakeholders as a model to save time in their group activity.

**Implications:** These findings suggest that a) the higher up in the organization one sits, the greater the emphasis on bottom line training results (i.e. the heavier weight given by Executives in the Outcomes Branch, b) the more skepticism about the current quality of the training. It would also appear that a balance of pay and status levels in stakeholder groups, especially if only one group is used, is advisable to counteract the natural biases of stakeholders from various levels in the organization.

**Evidence:** A comparison of prior and post utilities in both Tables 16 and 17 indicates that post utilities were consistently higher than prior utilities. This pattern is also reflected in the prior and post total impact scores also presented in Tables 16 and 17.

Two points will be offered regarding the significance of these findings. First, they suggest that program participants typically rate training higher than other stakeholders who are more removed from the training. Second, the findings add a third piece of information forming a pattern that suggests that the closer one is to the training experience itself, the more favorable one tends to view or evaluate the training.

The other pieces of this trend are: 1) Executive stakeholders tend to evaluate training lower than Mid-Management stakeholders, and 2) Mid-Management stakeholders tend to evaluate training lower than program participants.

Thus, a hierarchical pattern appears to exist and suggests that proximity to the training program has a strong, positive influence on how favorable one tends to view and evaluate that experience.

**Implications:** These findings follow both logically and empirically from what one might have expected regarding the favorable bias of training participants. It is quite plausible by logical grounds alone, for example, that training recipients of a generally favorable course would have favorable, pro-training attitudes which would influence their responses to surveys. Decision makers, in contrast, are more removed from the experience and would have more of a neutral attitude. The empirical basis for expecting these results would come from the literature on training evaluation which led to the discovery of the term "halo effect", referring to the positive bias held by students who are



surveyed immediately after taking an entertaining (if not effective) training course. This research also led to the reference to end of course feedback surveys as "smile sheets", alluding to the artificial inflation of student ratings of training attributed to the popularity of the instructor, and to the natural high that often occurs at the conclusion of training. Both sources of evidence, the logical and the empirical, would lead one to predict in the direction of the present findings.

**Implications:** The practical significance of this finding is simply that training evaluators using MAUT can expect a pro-training bias from participant level stakeholders, and that this bias will fade as one moves up the scale of stakeholders more removed from the actual training experience. This suggests, in turn, that various levels of stakeholders need to be involved in MAUT evaluations to get a balanced perspective.

## CHAPTER V

### SUMMARY, CONCLUSIONS, and RECOMMENDATIONS

#### Summary of Purpose

The present study was undertaken to learn more about MAUT as a method for the evaluation of training. An actual organizational training environment was chosen to apply MAUT and observe the strengths, weaknesses, successes, failures and conditions under which MAUT is appropriate or inappropriate. In addition, the investigation was designed to discover how to apply MAUT for training evaluation in ways that will maximize its effectiveness as an evaluation tool.

The approach consisted of applying MAUT to the evaluation of five training courses in the training department of a medium size city-government organization. Stakeholders were selected to represent various levels of management. Two stakeholder groups were used to develop separate sets of evaluation criteria, called MAUT attributes, prior to the evaluation. Using the steps of the MAUT process defined by Edwards et al. (1982), attributes were ranked and weighted, attribute indicators were developed, data were collected to provide measures of the indicators, and utility indices were derived. Finally, results were summarized, fed back to the stakeholders, and reactions and evaluations of the process were solicited. The specific categories of interest were: 1) Stakeholder Availability/Time Commitment; 2) Difficulty of Activities; 3) Use of Surveys; 4) Group Dynamics; 5) Organizational Politics; 6) Utilization of Results; 7) Accuracy/Credibility of Results; 8) Reaching Consensus; 9) Miscellaneous.

The investigation of MAUT is important to the professional training evaluation community because there has historically been a lack of quality evaluation in training organizations, and the systematic evaluation of human resource development programs for program

improvement is recognized as the best way to insure quality, desired impact, maximize efficiency, and control costs.

The relatively expensive cost of training versus other forms of communication justifies the cost of evaluation. However, the cost of experimental and quasi-experimental designs for impact evaluation, both in time and in obtaining the required expertise, plus the additional problems associated with these classical approaches (see Chapter II), suggest that other evaluation models such as MAUT are worth investigating as possible alternatives.

It was the thesis of this study that if MAUT could be successful in an actual training environment, it would provide training evaluators and training administrators with a more useful evaluation tool which could lead to either a greater use and practice of training evaluation, or to more useful evaluation in places where some evaluation was already taking place. But first, more needed to be learned about MAUT in an actual training environment.

In addition to the specific research questions, the literature on MAUT revealed that certain benefits are expected to be realized in the application of MAUT to evaluation problems. From the literature on the theory of MAUT, expectations (not quite hypotheses) were extracted about how MAUT could be useful. Before summarizing what was learned about MAUT in the present study, the list of the benefits/expectations of MAUT will be reiterated.

- 1) MAUT evaluations will make a difference in the decisions that get made about the programs evaluated;
- 2) Decision makers will view the MAUT technique/approach as appropriate for assessing the strengths and weaknesses of the programs in relevant dimensions of value;
- 3) A single decision maker is rarely the case in programs; therefore, MAUT will be effective at explicating, comparing, possibly reconciling, and

ultimately aggregating the values of groups in conflict about the desired value attributes of programs; 4) Decision makers will find the relatively simple statistics, the large amount of detail, and the quantitative data relevant to their values; and, they will find the reports useful and appealing, albeit potentially thick and heavy; 5) The decision makers will find the method easy to communicate and useful in situations where time is short and decision makers are numerous and busy; and, 6) The social psychology of small groups and group process effects will be evident in the required stakeholder group activities, such as the development and weighting of MAUT attributes.

#### Summary of Results

The study found that stakeholder time was very limited. Individual stakeholder interviews were easier to obtain, but very useful prior to group activities. These findings were quite similar to findings cited by Novak (1985), Miller (1985) and Curlette (1991). MAUT stakeholder activities took about ten hours to twelve hours. Stakeholder conceptual and communication skills are more crucial to the success of MAUT than was anticipated. Stakeholder participation and involvement in group discussion about training in their organizations was found to be positive and was cited as a strength of MAUT.

Stakeholders did identify some weaknesses in MAUT, such as complexity in the MAUT calculations, information overload, the time commitment, and the overall length of time needed for the study. This outcome was quite similar to Novak (1985) who noted that data in MAUT can become complex very quickly tending to overwhelm the participants.

The study took about nine months from the initial group workshops through the oral feedback to the stakeholder groups. It will be noted, however, that much of this time included the collection of survey data

from students entering training at staggered pre-scheduled intervals not under the control of the investigator, plus the desire to have post-course follow-up surveys distributed to students and supervisors no earlier than two months following the completion of a class.

Evidence indicated that, as in any small group exercise, factors such as organizational status, education, work experience, and expertise all play a role in the effectiveness of the group to debate relevant issues and reach consensus. MAUT workshop activities en route to attribute development and weighting, when done as stakeholder group exercises, create a natural forum for the dynamics of small group theory and process to be manifested. Organizational status and educational level were observed in this study as factors influencing communication among stakeholders in both groups.

The process of attribute development, ranking and weighting was found to be relatively straightforward and problem free, although it does require a degree of stakeholder ability to conceptualize training in terms of values and expectations. This ability level was not uniform among the stakeholder groups. It was also found that attribute development was not feasible or successful when done by survey.

Developing survey-type indicators for attributes was found to be relatively problem free. Further, surveys were found to be useful for collecting indicator measures, and for obtaining feedback from stakeholders about the MAUT process. Surveys were not very effective for developing or weighting attributes without direct guidance from the evaluator. Others, however, have reported the successful use of surveys to rate attributes (Garvill et al., 1992); Alpert et al., 1992; Teas and Wong, 1992; Sanbonmatsu et al., 1992; and, Curry et al., 1991).

This study successfully used the oral feedback of MAUT results to stakeholders and students in order to establish the credibility of the

data collected. It was also discovered that oral feedback provides a built-in error feedback process as a counter measure to control the relatively high error potential of the many arithmetic steps required in MAUT.

Other specific advantages of MAUT according to stakeholder feedback were the credibility of the results, the thoroughness of the results, the advantages for planning training, and the capability of MAUT to identify strengths and weaknesses of courses.

In terms of whether the MAUT results would make a difference in decisions about training, this study uncovered several relevant findings which collectively serve to inhibit the use of MAUT in decision making. First, decisions about training are made above the level of the stakeholders participating in this study. Second, top management makes intuitive decisions about training. Third, it was the stakeholders' opinion that MAUT recommendations would get quietly followed, but that top management would unlikely get fully briefed, partly due to the complexity and volume of the results. Traces of similar findings from actual education or training environments was also reported by Novak (1985), Miller (1985) and Curlette (1991).

A fourth line of evidence involved the Personnel Director's response to the imaginary memo and the recommendations it contained. His response to many of the recommendations was "Go for it!". On the other hand, as noted above, the real decision maker for training was the City Manager who, it was said, had ultimate decision power over training. According to the stakeholders, it was unlikely that the City Manager would be briefed or be influenced by the MAUT evaluations. Another problem cited was the large amount of rewriting/simplifying needed for the MAUT results to be communicated effectively to the City Manager and his/her immediate staff.

Related to the usefulness of MAUT was the question of whether MAUT generates simple statistics which are easy to understand by stakeholders. This study found that stakeholders complained about the complexity of the calculations rather than speaking positively about ease of understanding, simplicity, etc. It was also found that MAUT was perceived as difficult to communicate to and by stakeholders and to not be practical for use with small training staffs and evaluation resources. However, the present study was limited in the ability to test explicitly for these potential benefits of MAUT due to other added components (strategic value, etc.) and the up front decision to collect actual empirical data from students attending upcoming classes in the courses studied. In retrospect, had the investigator simply stopped after getting prior location measures from the seven stakeholder group members, not collected data from new students, and left out the strategic value and total impact components of the design, the stakeholders may very well have identified MAUT as being simple, easy to use and to communicate to others rather than complex.

As to whether MAUT was found useful for resolving value conflicts among stakeholder groups, this study found that the potential is certainly there for conflict resolution and for consensus building, given a certain level of stakeholder ability to deal with the cognitive components of those processes. In the present study, however, not a great amount of value conflict was found. Thus, it was not a great test of the conflict resolution aspect of MAUT.

As for unanticipated results, this study found that consensus in MAUT workshops was more easily reached in the executive group; that process and outcome branches appear to be natural categories for attributes in the training evaluation context; that prior location measures (LMs) were not correlated with post-LMs, suggesting that

decision makers and training participants view training merit/worth differently; that the MAUT results had validity to the extent that stakeholder rank ordering of general course utility was positively correlated with actual total utility scores; that stakeholder level in the organization had influence on the decision tree structure and weights of the attributes; that training participants typically rated training higher than did the stakeholder groups, and finally, that MAUT has the capacity to allow the evaluator to come quickly and closely in touch with the social and political context of the program under evaluation.

### Conclusions

Both theoretical and practical implications emerged from the study which are relevant for students and practitioners of social, educational, and training program evaluation.

**Theoretical:** In terms of the theory of program evaluation, it was clear that the MAUT approach allowed the evaluator to become deeply immersed in the politics and culture of the setting in which the evaluation was to take place. The ability of an evaluation method to enable this is immensely valuable in order to provide users with a broad-based evaluative assessment of programs being evaluated. In contrast to more narrowly focused, traditional impact evaluations, MAUT demonstrated the capacity to discover/produce not only quantitative measures of program merit and worth, but also to discover qualitative barriers, constraints, and other realities that have a bearing on the whole question of program impact. In so doing, MAUT as a process created a vivid reflection of the culture and setting that provided background and context for the programs and the evaluation.



For evaluation theorists, and for developers of theoretical models of program evaluation, a process such as MAUT is critically important, in the writer's view. Those barriers encountered which at the time are barriers to the successful and complete application of a sequential data collection and analysis process, are, when viewed from a different angle, the very organizational realities and cultural characteristics that define the context for understanding how and why a program operates as it does, and for better understanding the evaluation information collected. Thus, what may appear on the one hand to be an impediment to a MAUT evaluation, may in reality be a relevant organizational finding produced and/or exposed uniquely by MAUT rather than a contaminator of MAUT.

To illustrate briefly, is it really important that only four of the seven stakeholders in this study responded to the final survey which asked them to evaluate the MAUT process they had just been through? Or, is it much more important that MAUT exposed differences between mid-management and executive views about training and the value of evaluation, including the value of responding to evaluator questions about the MAUT process? Long term, it may be far more important for top management to be aware of specific differences between executive level and mid-management level views and values, than it is for the researcher to bemoan the lack of a 100% response rate to his survey.

**Stakeholder Participation:** For practitioners of training and educational evaluation, stakeholder time and stakeholder ability to deal with the attribute development and related activities are real issues in MAUT. Researchers or evaluators can expect stakeholders to be willing to participate in MAUT, and enjoy the activities and the intellectual challenge of the process. They will also enjoy the social aspects of the group tasks. Just getting together to discuss a common goal or

problem will be a positive experience.

Evaluators can also expect the dynamics of group process and the principles of small group theory to be relevant to, and evident in the activities of MAUT as in any other group activity or exercise.

**Role of the Evaluator:** Evaluators should also be advised that the role of the evaluator in MAUT requires a unique combination of skills not usually obtained in graduate programs in research and evaluation. One of the key insights derived from this study was that the role of the evaluator is not only critical for success, but that the role is much different by evaluators in more traditional methods. The success of MAUT can vary heavily depending on several factors related to the evaluator and what he/she brings to the study.

The evaluator in MAUT plays many roles. First, the evaluator must do important investigative work prior to involving stakeholders. One of critical outcomes is to gain rapport and build commitment for the study, and in particular support for the time and involvement of stakeholders. Another key outcome is the decisions about the number and level of stakeholders to be involved, and in decisions about stakeholder ability and selection. It is immensely important to assemble stakeholders who represent a variety of view points, but who also can contribute to the tasks involved in the MAUT process.

Second, the evaluator must act as group/team facilitator for the stakeholder groups. The evaluator must frame the evaluation problem, be expert in the MAUT steps, be sensitive to the information needs and political orientation of the stakeholders, and be analytically competent in determining when alternative strategies for getting needed information from stakeholders are warranted( e.g., switch to individual interview, incorporate surveys, etc.), or perhaps deciding when stakeholders need to be replaced.

Third, the evaluator must be a skilled interpreter and communicator of study results. The desired outcome in MAUT is that stakeholders and other decision makers be able to relate easily to the findings and are motivated to take necessary action. Accomplishing this outcome could require the evaluator to adopt a multi-media approach involving face to face briefings, leading joint evaluator/stakeholder data analysis sessions, or developing new/non-traditional ways of reporting results in written form.

The common link between these various roles is that the MAUT evaluator must communicate frequently and intensively with the program sponsors, stakeholders and any other parties who may be of assistance to the effort. In short, total immersion in the culture of the program is a good thing in MAUT!

In contrast to this role is the role of the dispassionate researcher/evaluator in the more classical approaches. In these approaches, the goal is usually to divorce the evaluator as much as possible from the evaluation method so as to avoid "experimenter" bias. Often, the emphasis is on the use of instruments which can be "objectively" scored and whose reliability and validity can be quantitatively measured.

Classical methods are more likely to treat evaluator judgement as a source of error, and therefore evaluator/method interaction is to be avoided. Even when interviewing by an evaluator is necessary, detailed interview protocols and other safeguards are all designed to control evaluator bias.

MAUT, in sharp contrast demands a high degree of interaction with the method of MAUT, and one could even argue that interaction not only is necessary, but is so important/essential that it defines the MAUT method. MAUT in fact draws its power from evaluator interaction

with the method, rather than loses power due to the interaction of evaluator with method.

The evaluator also requires certain skills, many of which are not taught/acquired in graduate programs in evaluation. MAUT places a premium on communications skills, knowledge of small group theory and group dynamics, the ability to persuade strong political voices of the need to be systematic and to communicate together about values and desired program attributes, and the ability to communicate MAUT results in ways that generate understanding and action. This audience is much different from the more scientific audience evaluators are trained to write to and for.

**Attributes:** The problems that emerge in the development of MAUT decision trees, attributes, ranks and weights are both logistical and interpersonal. That is, researchers/evaluators can expect to have some difficulty assembling stakeholders in organizations for sufficient lengths of time in order to properly brief them on the nature of the task and what is required. Another potential problem is the ability of stakeholders to deal with/relate to the cognitive requirements of such questions as "When all the data are in, what would you like to be able to say about program X?" It was found here that stakeholders occupying higher level jobs in the organization tend to have an easier time grasping the requirements of attribute development. Aside from logistical and cognitive ability issues, the attribute development, rank, and weight activities as recommended by Edwards et al. (1982) pose no particular problem for training evaluation. It will be noted, however, that researchers/evaluators can expect attribute branches in a training evaluation setting to parallel the Input-Process-Output systems model in terms of where the attributes are likely to cluster content-wise.

Stakeholder availability and ability problems found in the attribute development phase also apply to location measures and attribute indicators. To ease the reader's burden they will not be repeated here.

**Complexity:** In regard to the expected outcomes of the MAUT process in terms of credibility, usefulness of the results, and impact on decision making in the training environment, it is concluded that the MAUT procedure used in this study would need to be very much shortened, simplified, and the communication of results clarified and simplified for maximum results. The findings reported here indicated that in all likelihood the evaluation process used was more complex than the evaluation problem warranted. Part of this was surely due to the research nature of the study and the rather built-in constraint that in order to find out what would work one needed also to find out what didn't work, the result being more things tried than would be standard procedure in the applied environment.

Greater complexity was certainly added in this study due to the fact that five courses were being evaluated simultaneously, and that strategic values and course need were added to the equation for determining course impact. This, together with the complex (in the stakeholders' view) imaginary memorandum that was used as a written report of the MAUT results, resulted in feedback which was inconsistent with the expectations from the literature.

Although the writer acknowledges the negative reactions of the stakeholders, it must also be noted that the simplicity factor of MAUT can be manipulated, negated, or otherwise disrupted by the planned or inadvertent actions of the evaluator. In this study, for example, a glossary of terms could have been provided (but was not even considered necessary) as suggested by two of the stakeholders. This could have

made the report much simpler. In addition, the strategic value and total course impact add-ons could simply have been left out without serious loss or damage to the essential MAUT components, adding back in greater simplicity. The point here is that the negative feedback obtained in this study, while acknowledged, does not directly comment on the essential simplicity of MAUT for stakeholders and decision makers due to the obvious built-in complexity chosen by the investigator.

**Summary:** In summary, several findings relevant to the successful application of MAUT in the training environment were identified and discussed. The single overriding conclusion offered is that MAUT appears to offer much promise as an evaluation method in the training environment. In addition, the very nature of MAUT, in terms of a process that confronts stakeholder values and decision criteria directly, makes it superior in part to experimental and quasi-experimental designs, which typically are more narrowly focused and much less interested in (or capable of accommodating) stakeholder values. The present study exposed many of those organizational constraints and issues, and gained insight to the solution of many of them.

### **Recommendations**

In terms of the practical aspects of the application of MAUT, this investigation determined that the recommendations below appear warranted for researchers and evaluators in the training and educational evaluation environment. Additional recommendations appear in Appendix L in the form of a MAUT user's guide based on the findings of the present study:

- 1) **Advance Planning:** Invest a good deal of time prior to the study talking with potential stakeholder participants, program administrators, key decision makers, etc. This strategy is designed to gain support and commitment, begin orienting possible stakeholders to the MAUT activities, and

to orient the evaluator to the technical, social, and political context of the program.

- 2) Client Definition: Defining the client for the evaluation is critical. Evaluators should, to the extent possible, have a defined client for the evaluation and attempt strongly to locate and involve the real decision makers of the program in the evaluation.
- 3) Decision Maker(s) Contact: Interview the prime decision maker, even if he/she cannot participate in the evaluation with other stakeholders. Get his/her values and expectations about the program.
- 4) Stakeholder Selection: Get the highest level stakeholders possible for the attribute development and weighting tasks. This should ensure stakeholder ability to deal with the necessary activities.
- 5) Stakeholder Orientation: Interview all stakeholders individually prior to the group activities if possible. This will be a valuable orienting process for the stakeholders and an additional learning opportunity for the evaluator. Focus of the interview should be on prepping stakeholders for the coming activities, assuring them of the importance of their viewpoint and that they are not personally being evaluated, and on obtaining their initial thoughts on desired program attributes. The four orienting questions used in the present study (see Chapter III, Method) were very useful for this purpose.
- 6) MAUT Workshops: Conduct two four-hour sessions with the stakeholders; four hours for introduction, attribute development and weighting; four hours for prior location measures and indicator development. Try to get stakeholders to specify up front a standard up acceptability for the final grand Location Measure. The present study showed that stakeholders need/want a benchmark for evaluating utility values. This can be provided early in the evaluation by asking stakeholders to specify what percent of utility, below maximum, they would find acceptable.
- 7) Data Collection: Collecting needed empirical data, such as student or participant ratings, in small focus groups rather than by surveys. This will greatly speed up the process.
- 8) Stakeholder Feedback: Feedback results to stakeholders in a face to face briefing setting. Use only two or three tables to illustrate your points. Information overload is a large risk!
- 9) Final Report: Write a final report to management that is basically only an executive summary. Include observations about the organizational culture that are relevant to program operation as well as reporting the evaluation data per se.

- 10) Evaluation Team: Strong consideration should be given to using at least some of the stakeholders as formal evaluation team members to enhance stakeholder commitment of time and enthusiasm.
- 11) Evaluator Training: The role of the MAUT evaluator is unique in many ways detailed in this report. It requires training in several areas not normally provided by graduate programs in evaluation. Evaluators should obtain training in conducting focus group workshops or in group dynamics in addition to mastering the technical aspects of MAUT.

Appendix L contains an elaboration of these suggestions in the form of a MAUT Evaluators Guide for Educational and Training Evaluation.

In terms of recommendations for further research, the general recommendation is to test the practical suggestions noted above. Specifically, the following items are offered for further research to determine whether MAUT can yield greater acceptability among stakeholders than was obtained in this study (note: the reader will recall the findings which highlighted stakeholder complaints about complexity of MAUT results, for example) and have greater influence on decision making:

- 1) Research is needed which develops and tests strategies for involving "real decision makers" in the evaluation.
- 2) Research is needed which develops and tests ways for obtaining the highest level stakeholders to ensure stakeholder ability to work effectively through the MAUT process.
- 3) Research is needed to develop and test ways of identifying standards during the indicator development process. The issue here is providing a priori standards for attribute and overall total utility scores in order to provide a basis for decision making when MAUT is used for evaluation. That is, when MAUT is used for deciding between two options, that option which yields the largest overall utility value is usually the preferred option. But when MAUT is used to evaluate a single program, a priori standards of acceptability are needed, developed by stakeholders, to aid in determining whether the program has "acceptable utility" in terms of MAUT generated utility scores. Standards can be developed at both the attribute and overall levels.



- 4) Research is needed to develop and test the use of focus groups to collect data for use in developing location measures in lieu of surveys.
- 5) Research is needed to develop and test alternative ways of reporting MAUT results to management. The memo method used in the present study, designed to be concise and to the point, was still viewed as too complex by the stakeholders. The exploration of various alternative reporting formats for MAUT results would be worthwhile.
- 6) Research is needed to determine the effects of forming an evaluation team of stakeholders on MAUT results and the use of MAUT for decisions about programs. That is, a single group of six to eight stakeholders functioning as an evaluation team, something not done in the present study, appears worthwhile in order to gain greater commitment and lessen the problem of stakeholder availability. Team members would be involved with all phases of the MAUT evaluation, including oral and written reporting of results.

## REFERENCES

- Alpert, F. H., Kamins, M. A., & Graham, J. L. (1992). An examination of reseller buyer attitudes toward order brand entry. Journal of Marketing, 56(3), 25-37.
- Bakken, D., & Bernstein, A. (1982). A systematic approach to evaluation. Training and Development Journal, August.
- Ball, S., & Anderson, S. (1975). Practices in program evaluation: A survey and some case studies. Princeton, NJ, ETS.
- Baron, J. (1977). An exploration of the implication of the MAUT-decision theoretic model for summative and formative evaluation and post-assessment organizational change. Paper presented at Annual Meeting of American Educational Research Association, April, 4-8.
- Blumenfeld, W. S. (1966). Attitude change as a criterion in training. Training and Development Journal, October, 26-32.
- Bogdan, R. C., & Biklen, S. K. (1982). Qualitative research for education: An introduction to theory and methods. Allyn and Bacon.
- Brandenburg, D. (1982, August). Training evaluation: What's the current status?" Training and Development Journal, 36, 14-19.
- Brandenburg, D., & Smith, M. E. (1986). Evaluation of corporate training programs. ERIC Clearinghouse on Tests, Measurement, and Evaluation, ETX. Princeton, NJ.
- Brewster, C. (1980). Evaluation of management training: A focus on change. In J. Beck & C. Cox (Eds.). Advances in management education. Wiley and Sons.
- Brinkerhoff, R. O. (1981). Making evaluation more useful." Training and Development Journal. December, 66-70.
- Brinkerhoff, R. O. (1987). Achieving results from training. Jossey-Bass.
- Brumet, M. E. (1976). Applying Bayesian statistics to educational evaluation (Theoretical Paper No. 62). University of Wisconsin: Research and Development Center for Cognitive Learning (ERIC Document Reproduction Service No. ED 134 609)
- Campbell, D. T. & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Campbell, J. P., Dunnette, M. P., Lawler, E. E. IL, & Weick K. E., Jr. (1970). Managerial behavior and performance effectiveness. McGraw-Hill.
- Carrol, J. D. & de Soete, G. (1991). Toward a new paradigm for the study of multi-attribute choice behavior: Spatial and discrete modeling of pairwise preferences. American Psychologist, 46(4), 342-351.

- Catalanello, R. E., & Kirkpatrick, D. L. (1968, May). Evaluating training programs - the 'state-of-the-art'. Training and Development Journal, 22, 2-9.
- Clement, R. W. (1981). Evaluating the effectiveness of management training: Progress during the 1970s and prospects for the 1980s. Human Resources Management. Winter.
- Cline, M. G. (1988). In consultation with the writer.
- Cook, T. D., & Campbell, D. T. (1975). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.
- Curlette, W. L. (1991). An evaluation model to select an integrated learning system in a large suburban school district. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1991. (ERIC Document Reproduction Service No. ED 334 236)
- Curry, D. J., Menasco, M. B., & Van-Ark, J. W. (1991). Multi-attribute dyadic choice: Models and tests. Journal of Marketing Research, 28(3), 259-267.
- Demeuse, K. P., & Liebowitz, S. J. (1981, September). An empirical analysis of team-building research. Group Organizational Studies, 6, 357-378.
- Dubin, S. S., Mezack, M., & Neidig, R. (1974). Improving the evaluation of management development programs. Training and Development Journal. June, 42-46.
- Dyer, J. S., Fishburn, P. C., Steuer, R. E., Wallenius, J., & Zionts, S. (1992). Multiple Criteria Decision Making, Multiattribute Utility theory: The next ten years. Management Science, 38(5), 645-654.
- Edwards, W., & Guttentag, M. (1975). Experiments and evaluations: A re-examination. In C. A. Bennett & A. A. Lumsdaine (Eds.). Evaluation and experiment: Some critical issues in assessing social program. New York: Academic Press.
- Edwards, W., Guttentag, M., & Snapper, K. (1975). A decision theoretic approach to evaluation research. In L. Struening & M. Guttentag (Eds.). Handbook of evaluation research (Vol. 1). Beverly Hills, CA: Sage.
- Edwards, W., & Newman, J. R. (1982). Multi-attribute evaluation. Quantitative Applications in the Social Sciences Series, Beverly Hills, CA: Sage.
- Einhorn, H. J., & McCoach, W. (1977). A simple multiattribute-utility procedure for evaluation. Behavioral Science, 22, 270-282.
- Elsbree, A. R., & Howe, C. (1977). An evaluation of training in three acts. Training and Development Journal, July.

- Garre, P. P. (1992). Multiattribute Utility Theory in decision-making. Nursing Management, 23(5), 33-35.
- Garvill, J., Garling, T., Lindberg, E., & Montgomery, H. (1992). Economic and non-economic motives for residential preferences and choices. Journal of Economic Psychology, 13(1), 39-56.
- Geroy, G. D., & Wright, P. C. (1988). Evaluation research: A pragmatic, program-focused research strategy for decision-makers. Performance Improvement Quarterly, 1, No. 3.
- Goldstein, I. L. (1974). Training: program development and evaluation. Monterey, CA: Brooks/Cole
- Goldstein, I. L. (1980). Training in Work Organizations. Annual review of psychology (Vol. 31, pp. 229-72).
- Goldstein, I. L. (1986). Training in organizations: needs assessment, development, and evaluation (2d ed.). Monterey, CA: Brooks/Cole.
- Grider, D. T., & Capps, C. J. (1990). Training evaluation: Prescription for improvement. Business, 40(1), 20-24.
- Guba, E. G., & Lincoln, Y. S. (1981). Effective evaluation. San Francisco: Jossey-Bass.
- Hamblin, A. C. (1974). Evaluation and control of training. McGraw-Hill.
- Hawthorne, E. (1985). Conceptual foundations of evaluation research: A case study of corporate management education. Unpublished dissertation. (University Microfilms No. 8600453).
- Hinrichs, J. R. (1976). Personnel training. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.
- Hogarth, R. H. (1979). Evaluating with validity. Beverly Hills: Sage.
- Joint Committee on Standards for Educational Evaluation (1981). Standards for evaluation of educational programs and materials. New York: McGraw-Hill.
- Kirkpatrick, D. L. (1959, 1960). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 13, 3-9, 21-26; 14, 13-18, 28-32.
- Kirkpatrick, D. L. (1977). Evaluating training programs: Evidence vs. proof. Training and Development Journal, November.
- Kohli, R., & Mahajan, V. (1991). A reservation-price model for optimal pricing of multiattribute products in conjoint analysis. Journal of Marketing Research, 28(3), 347-354.
- Komaki, J. (1977). Alternative evaluation strategies in work settings: Reversal and multiple base-line designs. J. Organizational Behavior Management, 1, 53-77.

- Kopalle, P., & Hoffman, D. (1992). Generalizing the sensitivity conditions in an overall index of product quality. Journal of Consumer Research, 18(4), 530-540.
- Larson, R. C., & Kaplan, E. H. (1981). Decision oriented approaches to program evaluation. New directions for program evaluation (Vol. 10, pp. 49-68).
- Latham, G. P. (1988). Human resource training and development. Annual review of psychology (Vol. 39, pp. 545-582).
- Lipsey, M. W. (1988, Nov.). Practice and malpractice in evaluation research. Evaluation Practice, 9, p. 4.
- Miller, A. (1985). Gifted and talented program evaluation: A multiattribute utility approach. Evaluation report for Elliott County School Corporation, Morgan County Corporation, Rowan County School Corporation. (ERIC Document Reproduction Service No. ED 283 854)
- Nadler, L. (1982). Designing training programs - The critical events model: Addison-Wesley.
- Nevo, D. (1986). The conceptualization of educational evaluation: An analytical review of the literature. Evaluation: Philadelphia: Palmer Press.
- Novak, C. D. (1985). A systematic procedure for setting evaluation team priorities. Paper presented at the annual meeting of the American Education Research Association, Chicago, 1985. (ERIC Document Reproduction Service No. ED267 117)
- Park, Byung-Yang (1979). The development and application of a decision-theoretic evaluation method. Unpublished dissertation. (University Microfilms No. 8004825).
- Patton, M. Q. (1980). Qualitative evaluation methods. Beverly Hills, CA: Sage.
- Rossi, P. H., & Williams, W. (Eds.). (1972). Evaluating social programs. Seminar Press.
- Rummler, G. A., & Brethower, K. S. (1979, May). Evaluating Training. Training and Development Journal.
- Rutman, L. (1980). Planning useful evaluations. Beverly Hills, CA: Sage.
- Sanbonmatsu, D., Kardes, F., & Herr, P. (1992). The role of prior knowledge and missing information in multiattribute evaluation. Organizational Behavior and Human Decision Processes, 51(1), 76-91.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.). Perspectives of curriculum education. Chicago: Rand McNally.

- Shoemaker, P. J., & Waid, C. C. (1988). A probabilistic dominance measure for binary choices: Analytic aspects of a multiattribute random weights model. Journal of Mathematical Psychology, 32(2), 169-191.
- Smith, M. E. (1984, April). Trends in training evaluation. Paper presented at the NSPI Paper Presentation, Atlanta.
- Snyder, R. A., Raben, C. S., & Farr, J. L. (1980). A model for the systematic evaluation of human resources development programs. Academy of Management Review, 5(3), 431-444.
- Stake, R. E. (1967). Toward a technology for the evaluation of educational programs. In R. W. Tyler, R. M. Gagne, & M. A. Scriven (Eds.). Perspectives of curriculum education. Chicago: Rand McNally.
- Stufflebeam, D. I., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. I., Merriman, H. O., & Provus, M. M. (1971). Educational evaluation and decision-making. Itasca, IL: F. E. Peacock.
- Tannenbaum, S., & Yukl, G. (1992). Training and development in work organizations. In Annual Review of Psychology, 43, 299-342.
- Teas, R. K., & Wong, J. K. (1992). Item content and the stability of entity-based and attribute-based multiattribute scaling methods. Journal of Consumer Research, 18(4), 536-545.
- Thomas, J. B., McDaniel, R. R., & Dooris, M. J. (1989). Strategic issue analysis: NGT + decision analysis for resolving strategic issues. Journal of Applied Behavioral Science, 25(2), 189-200.
- Troutt, M. D. (1991). On some multiattribute value function generalizations of the EOQ model in the context of personal inventory decisions. Theory and Decision, 30(2), 95-107
- Tyler, R. W. (1950). Basic principles of curriculum and instruction. Chicago, IL: University of Chicago Press.
- Vernon, D. R. (1985). Two approaches to evaluation: A comparison of quasi-experimental and bayesian, MAUT designs. Unpublished dissertation. (University Microfilms No. 8521004)
- Weiss, C. H. (1972a). Evaluation research methods of assessing program effectiveness. Prentice-Hall.
- Weiss, C. H. (1972b). Evaluating educational and social action programs. In C. H. Weiss (Ed.), Evaluating action programs, readings in social action, and education (pp. 1-25). Boston: Allyn and Bacon.
- Wexley, K. N. (1984). Personnel training. Annual review of psychology (Vol. 35, pp. 519-51).
- Zenger, J. H., & Hargis, K. (1982, January). Assessing training results: It's time to take the plunge. Training and Development Journal, 11-16.

## **Appendices**

## Appendix A

TO: Executive Stakeholder Group

Attached are the edited samples of desired program attributes.

The next step is to develop a decision value tree such as the one on page 1 of last week's handout (attached).

Here are the steps of that process. Please use whatever worksheets you need to do the work, then prepare a final cleaned-up version for discussion with me and other members of the team. I will be calling you about September 3 to see how you are doing.

Step 1 - Using the attached samples ONLY AS A GUIDE AND/OR REMINDER OF WHAT HAS BEEN SAID THUS FAR, develop your own desired course attributes by answering the question: "After all the data are in, what would you like to be able to say about the training programs under review?" Use short phrases or sentences, as shown on page 1 of previous handout. I'm looking for a minimum of 8 and a maximum of 15 attributes.

Step 2 - Group your attributes, if possible, by their common theme. Label the group as if it were a branch of a tree, with the attributes beneath each branch serving as "twigs."

Step 3 - Arrange the twigs within each branch in order of importance, with the most important twig listed first, followed by the next most important, etc.

Step 4 - Within each branch assign the LEAST important twig the value ten (10). Then, assign the next least important twig a value that indicates its importance relative to the first twig rated. Thus, if the next least important twig is rated a 20, it will mean that it is twice as important as the first one rated; if rated 30 it is three times as important, and so on. Then move up to the third least important twig in the branch and rate it, and so on, all ratings within the branch being rated relative to the one rated 10.

Step 5 - Move on to the next branch and start over by giving the least important twig a value of ten (10), and so on, until you have completed all the twigs within all the branches.

NOTE: The most important twig within a branch does not have to be limited to a rating of 100. The only constraint is that the least important twig within the branch be rated a ten (10).

Step 6 - For each twig in every branch, assign six (6) ratings (one for each of our six courses--course list also attached). A rating from 0 to 100 will represent the extent to which you feel each course possesses the desired attribute. A 100 represents the maximum possession of the attribute, and 0



represents zero possession of the attribute. This is to be a "gut level" feeling you have about each course. Don't worry about precision or accuracy. We are trying to get you to put a value to your "hunch" about each course before we go out and collect some real data.

THANKS FOR YOUR CONTINUED PARTICIPATION. FEEL FREE TO WORK TOGETHER TOWARD CONSENSUS AFTER FIRST WORKING INDEPENDENTLY!!

John Milatzo

Date: 8/27/91

**SAMPLE DESIRED SYSTEM ATTRIBUTES \***

<u>Branch/Twig (Branch Weight)</u>	<u>Standardized Within Branch Weight</u>	<u>Total Weight</u>
• User Perspective (.23)		
• Enables career planning	.36	.08
• Measures interests, skills, values	.25	.06
• Enables better understanding of self, career goals, options, etc.	.18	.04
• Useable information obtained	.14	.03
• Convenient to use	.07	.02
• EDS Perspective (.35)		
• Provides better understanding of counselor role	.26	.09
• Helps prevent burnout	.21	.07
• System viewed as advantageous	.17	.06
• More people reached	.14	.05
• Useful counseling tool	.10	.04
• System is superior to the competitor	.07	.02
• System can/has been sold to EDS' superiors	.05	.02
• MED Perspective (.12)		
• System is supported by MEDs	.12	.12
• Division Perspective (.29)		
• System will be supported by Divisions	.29	.29

\* Example taken from a MAUT evaluation of a computerized career development program.

**Potential "Desired" Course Attributes**

The program creates favorable reactions that are useful in motivating, retaining, and recruiting greatly needed employees.

The program enhances the value of participating personnel by increasing their capacity to perform in future roles and situations not related to their current jobs.

The program increases important job performance; training graduates perform important tasks with greatly enhanced productivity.

Learning is assured through proper delivery level and style.

Training leads to use in one's personal life as well as on-the-job.

Student participation and discussion is part of the method; adult learners are kept interested.

Instructor(s) encourages feedback and student interaction.

Students learn from the other students as well as from the instructor.

Training makes the job easier for the students.

Should move at comfortable pace.

Instructor should know material well, know audience well, and tailor material wherever possible.

There is a legitimate need for the course.

Instructor takes a personal approach and uses many examples from his/her own personal experience to illustrate points.

Has a (measurable) impact on the work force.

Training goals and objectives are met; course is faithful to its stated purpose.

Course is well received by students; students enjoy the experience.

There is a lively, energetic atmosphere in class.

Instructor inspires and motivates students.

Course includes useful handouts, books, etc. which help the student after the course.

A course should make you think; you should have homework; you should be challenged.

Empowerment of employee to leave training with a sense of being a stronger person in the work place.

Empowerment of having a skill applicable to whatever employee does (in and out of work).

High cost benefit--within ten days I would expect more from the employee.

Return to work place with a text/manual/or other important job aid.

High benefit/cost ratio (delivered in one day as opposed to one week).

Content coincides with Department Head needs.

Employee returns to the job with a heightened enthusiasm for his/her job.

Broader perception of job fit to the organization.

**Appendix B****EXECUTIVE DECISION VALUE TREE****PROCESS ATTRIBUTES**

**ATTRIBUTE:** Content aims at important needs.

**INDICATORS:**

3. To what extent does the above named employee have a direct job performance need for the above named course? (Pre-Training Survey/Supervisor)

1	2	3	4	5	6	7
Small Extent						Large Extent

2. To what extent do you have a direct job performance need for the above named course? (Pre-Training Survey/Student)

1	2	3	4	5	6	7
Small Extent						Large Extent

12. The course is related to important aspects of my job. (Training Feedback Survey)

1	2	3	4	5	6	7
Strongly Disagree --			Neither Agree nor Disagree --		Strongly Agree	

ATTRIBUTE: Learning is insured; knowledge or skill is gained.

INDICATORS:

1. My knowledge/skill level in the subject matter at the conclusion of this course is: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low						Very High

15. In retrospect, my knowledge/skill level before taking this class was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low						Very High

3. I would rate my employee's present skill level in the subject as: (Post-Training Survey/Supervisor)

1	2	3	4	5	6	7
Very Low						Very High

4. Compared to my employee's knowledge/skill level now, I would rate his/her knowledge/skill level before the training as: (Post-Training Survey/Supervisor)

1	2	3	4	5	6	7
Very Low						Very High

ATTRIBUTE: High benefit/cost ratio.

INDICATORS:

3. My time is valuable so I consider the time I spent in this class to be: (Training Feedback Survey)

1	2	3	4	5	6	7
Not Well Spent						Very Well Spent

ATTRIBUTE: Effective, competent instructor.

INDICATORS:

4. The instructor's knowledge of the subject matter was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low			Adequate			Very High

5. The instructor's effectiveness in facilitating a learning experience was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low						Very High



ATTRIBUTE: Well received/enjoyed by the students.

INDICATORS:

6. My reaction and enjoyment of this course was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Negative			Adequate			Very Positive

ATTRIBUTE: Useful text, handouts, etc.

INDICATORS:

7. The quality of the course material (handouts, texts, etc. - if applicable) was: (Training Feedback Survey)

1	2	3	4	5	6	7
Poor			Adequate			Excellent

8. The quality of the audio/visual aids (if applicable) was: (Training Feedback Survey)

1	2	3	4	5	6	7
Poor			Adequate			Excellent

ATTRIBUTE: Learning from other students possible.

INDICATORS:

9. To what degree was the course presented so that the students could interact with and learn from each other? (Training Feedback Survey)

1	2	3	4	5	6	7
Low Degree			Adequate			High Degree

ATTRIBUTE: Teaching methods matches material and audience needs.

INDICATORS:

10. The teaching method in this course was well suited to the material and to audience needs. (Training Feedback Survey)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

ATTRIBUTE: Lively, energetic atmosphere.

INDICATORS:

11. The atmosphere in this course was highly energetic and positive.  
(Training Feedback Survey)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

OUTCOME ATTRIBUTES

ATTRIBUTE: Increases current job performance.

## INDICATORS:

1. Since my employee attended this training, I have noticed a positive change in his/her job performance. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

6. Since attending this training, my employee has shown greater motivation on the job. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

7. Since attending this training, my employee has shown a more positive attitude. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

8. Since attending this training, my employee has improved his/her job performance \_\_\_\_\_ percent in tasks and duties related to this training: (Post-Training Survey/Supervisor)

0%   10%   20%   30%   40%   50%   60%   70%   80%   90%   100%

1. Since attending this training, there has been a positive change in my job performance. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

4. Since attending this training, I have experienced greater motivation on the job. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

5. Since attending this training, I have experienced a more positive attitude on the job. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

6. Since attending this training, I have improved my job performance \_\_\_\_\_ percent in tasks and duties related to this training. (Post-Training Survey/Student)

0%   20%   30%   40%   50%   60%   70%   80%   90%   100%

ATTRIBUTE: Provides useful job aids/tools.

INDICATORS:

7. The quality of the course material (handouts, texts, etc. - if applicable) was: (Training Feedback Survey)

1	2	3	4	5	6	7
Poor			Adequate			Excellent

8. The quality of the audio/visual aids (if applicable) was: (Training Feedback Survey)

1	2	3	4	5	6	7
Poor			Adequate			Excellent

ATTRIBUTE: Students empowered with heightened self-esteem, confidence, competence, etc.

INDICATORS:

5. I have observed a greater degree of comfort or confidence in my employee's performance in certain situations or tasks since he/she has attended this training. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

3. I have experienced a greater degree of comfort or confidence in my job performance in certain situations or tasks since attending this training. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree



ATTRIBUTE: Job tasks made easier.

INDICATORS:

8. This training has enabled me to perform certain tasks or duties more efficiently. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

ATTRIBUTE: Contributes to professional growth.

INDICATORS:

9. I believe this training will contribute to my professional growth.  
(Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

ATTRIBUTE: Practical application beyond the job.

INDICATORS:

7. I have used aspects of what I learned in this training in my private life. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

## Appendix C

### MID-MANAGEMENT DECISION VALUE TREE

#### PROCESS ATTRIBUTES

ATTRIBUTE: Content related to needs.

#### INDICATORS:

3. To what extent does the above named employee have a direct job performance need for the above named course? (Pre-Training Survey/Supervisor)

1	2	3	4	5	6	7
Small Extent						Large Extent

2. To what extent do you have a direct job performance need for the above named course? (Pre-Training Survey/Student)

1	2	3	4	5	6	7
Small Extent						Large Extent

12. The course is related to important aspects of my job. (Training Feedback Survey)

1	2	3	4	5	6	7
Strongly Disagree	--	Neither Agree nor Disagree	--	Strongly Agree		

ATTRIBUTE: Learning is insured (through good design).

INDICATORS:

1. My knowledge/skill level in the subject matter at the conclusion of this course is: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low						Very High

15. In retrospect, my knowledge/skill level before taking this class was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low						Very High

3. I would rate my employee's present skill level in the subject as: (Post-Training Survey/Supervisor)

1	2	3	4	5	6	7
Very Low						Very High

4. Compared to my employee's knowledge/skill level now, I would rate his/her knowledge/skill level before the training as: (Post-Training Survey/Supervisor)

1	2	3	4	5	6	7
Very Low						Very High

ATTRIBUTE: Training is participative; interaction encouraged.

INDICATORS:

9. To what degree was the course presented so that the students could interact with and learn from each other? (Training Feedback Survey)

1	2	3	4	5	6	7
Low Degree			Adequate			High Degree

ATTRIBUTE: Instructor inspires, motivates, creates energetic atmosphere.

INDICATORS:

4. The instructor's knowledge of the subject matter was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low			Adequate			Very High

5. The instructor's effectiveness in facilitating a learning experience was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Low						Very High

13. The instructor inspired and motivated students. (Training Feedback Survey)

1	2	3	4	5	6	7
Strongly Disagree		Neither Agree nor Disagree				Strongly Agree

ATTRIBUTE: Useful texts, handouts, etc.

INDICATORS:

7. The quality of the course material (handouts, texts, etc. - if applicable) was: (Training Feedback Survey)

1	2	3	4	5	6	7
Poor			Adequate			Excellent

8. The quality of the audio/visual aids (if applicable) was: (Training Feedback Survey)

1	2	3	4	5	6	7
Poor			Adequate			Excellent



ATTRIBUTE: Instructor takes personal approach; illustrates points with personal examples.

INDICATORS:

14. The instructor took a personal approach, used examples from real life, etc. (Training Feedback Survey)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

OUTCOME ATTRIBUTES

ATTRIBUTE: Training improves job performance.

## INDICATORS:

1. Since my employee attended this training, I have noticed a positive change in his/her job performance. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

7. Since attending this training, my employee has shown a more positive attitude. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

8. Since attending this training, my employee has improved his/her job performance \_\_\_\_\_ percent in tasks and duties related to this training: (Post-Training Survey/Supervisor)

0%   10%   20%   30%   40%   50%   60%   70%   80%   90%   100%

1. Since attending this training, there has been a positive change in my job performance. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

5. Since attending this training, I have experienced a more positive attitude on the job. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

6. Since attending this training, I have improved my job performance \_\_\_\_\_ percent in tasks and duties related to this training. (Post-Training Survey/Student)

0%   20%   30%   40%   50%   60%   70%   80%   90%   100%

ATTRIBUTE: Goals & objectives met; faithful to stated purpose.

INDICATORS:

2. My objectives and expectations in sponsoring this employee for training appear to have been satisfied. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

2. Speaking only for myself, I would say that this training satisfied its stated objectives. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

ATTRIBUTE: Student confidence enhanced.

INDICATORS:

5. I have observed a greater degree of comfort or confidence in my employee's performance in certain situations or tasks since he/she has attended this training. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

3. I have experienced a greater degree of comfort or confidence in my job performance in certain situations or tasks since attending this training. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

ATTRIBUTE: Positive student reactions; motivated employees.

INDICATORS:

6. My reaction to and enjoyment of this course was: (Training Feedback Survey)

1	2	3	4	5	6	7
Very Negative			Adequate			Very Positive

ATTRIBUTE: Training used in one's personal life.

INDICATORS:

7. I have used aspects of what I learned in this training in my private life. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

ATTRIBUTE: Employee motivation increased.

INDICATORS:

6. Since attending this training, my employee has shown greater motivation on the job. (Post-Training Survey/Supervisor)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

4. Since attending this training, I have experienced greater motivation on the job. (Post-Training Survey/Student)

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

**Appendix D**

**PRE-TRAINING SURVEY/SUPERVISOR**

Manager/supervisor of \_\_\_\_\_  
Employee/training participant's name

Your Name \_\_\_\_\_ Date: \_\_\_\_\_

-----  
The above named employee has applied for enrollment in the course  
\_\_\_\_\_. This course will be offered on  
\_\_\_\_\_.

This survey is designed to gather critical information from the employee's supervisor/manager prior to the start of the training. The information will be used along with other information to monitor training needs among employees and to evaluate and improve training programs.

Please answer the questions by filling in the blanks or circling the number that best describes your thoughts and opinions.

- 1. What is your best guesstimate of the percentage of the work force who have a direct job performance need for the above named course? \_\_\_\_\_ %
- 2. What percentage of the employees who you supervise do you feel has a direct job performance need for the above named course? \_\_\_\_\_ %
- 3. To what extent does the above named employee have a direct job performance need for the above named course?  

1	2	3	4	5	6	7
Small Extent						Large Extent
- 4. I evaluate my (above named) employee's knowledge/skill level prior to attending this training as:  

1	2	3	4	5	6	7
Very Low						Very High
- 5. I expect attendance at this training to raise my (above named) employee's knowledge/skill level to:  

1	2	3	4	5	6	7
Very Low						Very High
- 6. What percentage of the total working time would you estimate the above named employee spends on tasks that will be affected by this course? \_\_\_\_\_ %

Thank you very much for completing this survey. Please return it to Personnel Services Dept., Box 52.



## Appendix E

## PRE-TRAINING SURVEY/STUDENT

Training participant's name: \_\_\_\_\_

Date: \_\_\_\_\_ Course Title: \_\_\_\_\_

-----

This survey is designed to gather critical information prior to the start of the training. The information will be used along with other information to monitor training needs among employees and to evaluate and improve training programs.

Please answer the questions by filling in the blanks or circling the number that best describes your thoughts and opinions.

1. What percentage of the work force do you feel has a direct job performance need for the above named course? \_\_\_\_\_ %

2. To what extent do you have a direct job performance need for the above named course?

1	2	3	4	5	6	7
Small Extent						Large Extent

3. I evaluate my knowledge/skill level prior to attending this training as:

1	2	3	4	5	6	7
Very Low						Very High

4. I expect attendance at this training to raise my knowledge/skill level to:

1	2	3	4	5	6	7
Very Low						Very High

Thank you very much for completing this survey. Please return it to the instructor.

**Appendix F****TRAINING FEEDBACK SURVEY**

NAME: \_\_\_\_\_ DATE: \_\_\_\_\_

COURSE TITLE: \_\_\_\_\_

INSTRUCTIONS: Please circle the number that best describes your thoughts and opinions about this course.

1. My knowledge/skill level in the subject matter at the conclusion of this course is:

1	2	3	4	5	6	7
Very Low						Very High

---

2. Compared to my knowledge/skill level now, I expected my level after the course to be:

1	2	3	4	5	6	7
Very Low						Very High

---

3. My time is valuable so I consider the time I spent in this class to be:

1	2	3	4	5	6	7
Not Well Spent						Very Well Spent

---

4. The instructor's knowledge of the subject matter was:

1	2	3	4	5	6	7
Very Low			Adequate			Very High

---

5. The instructor's effectiveness in facilitating a learning experience was:

1	2	3	4	5	6	7
Very Low						Very High

---

6. My reaction to and enjoyment of this course was:

1	2	3	4	5	6	7
Very Negative			Adequate			Very Positive

---

7. The quality of the course material (handouts, texts, etc. - if applicable) was:

1	2	3	4	5	6	7
Poor			Adequate			Excellent

(Cont'd.)

8. The quality of the audio/visual aids (if applicable) was:

1	2	3	4	5	6	7
Poor			Adequate			Excellent

---

9. To what degree was the course presented so that the students could interact with and learn from each other?

1	2	3	4	5	6	7
Low Degree			Adequate			High Degree

---

10. The teaching method in this course was well suited to the material and to audience needs.

1	2	3	4	5	6	7	
Strongly Disagree	--	Neither	Agree	nor	Disagree	--	Strongly Agree

---

11. The atmosphere in this course was highly energetic and positive.

1	2	3	4	5	6	7	
Strongly Disagree	--	Neither	Agree	nor	Disagree	--	Strongly Agree

---

12. The course is related to important aspects of my job.

1	2	3	4	5	6	7	
Strongly Disagree	--	Neither	Agree	nor	Disagree	--	Strongly Agree

---

13. The instructor inspired and motivated students.

1	2	3	4	5	6	7	
Strongly Disagree	--	Neither	Agree	nor	Disagree	--	Strongly Agree

---

14. The instructor took a personal approach, used examples from real life, etc.

1	2	3	4	5	6	7	
Strongly Disagree	--	Neither	Agree	nor	Disagree	--	Strongly Agree

---

15. In retrospect, my knowledge/skill level before taking this class was:

1	2	3	4	5	6	7
Very Low						Very High

---

16. My overall opinion of this class is:

1	2	3	4	5
Poor				Excellent

## Appendix G

### POST-TRAINING SURVEY/STUDENT

Training Participant's name: \_\_\_\_\_

Date: \_\_\_\_\_

Course Title: \_\_\_\_\_

This survey is designed to gather critical information from you, a participant in the above named course, now that you are back on the job. This information will be used along with other information to monitor training needs among employees and to evaluate and improve training programs.

Please answer the questions by filling in the blank, or circling the number that best describes your thoughts and opinions.

1. Since attending this training, there has been a positive change in my job performance.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

2. Speaking only for myself, I would say that this training satisfied its stated objectives.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

3. I have experienced a greater degree of comfort or confidence in my job performance in certain situations or tasks since attending this training.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

4. Since attending this training, I have experienced greater motivation on the job.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

5. Since attending this training, I have experienced a more positive attitude on the job.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

6. Since attending this training, I have improved my job performance \_\_\_\_\_ percent in tasks and duties related to this training.

0%    20%    30%    40%    50%    60%    70%    80%    90%    100%

Continued

7. I have used aspects of what I learned in this training in my private life.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

8. This training has enabled me to perform certain tasks or duties more efficiently.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

9. I believe this training will contribute to my professional growth.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

10. What percentage of your total working time would you estimate you spend on tasks related to this course?

\_\_\_\_\_ %



7. Since attending this training, my employee has shown a more positive attitude.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

8. Since attending this training, my employee has improved his/her job performance \_\_\_\_\_ percent in tasks and duties related to this training:

0%    10%    20%    30%    40%    50%    60%    70%    80%    90%    100%

9. As a result of my employee's experience with this training, I will send additional employees to this training if the City offers it again.

1            2            3            4            5            6            7  
Strongly Disagree -- Neither Agree nor Disagree -- Strongly Agree

## Appendix I

### Summaries from Student Interviews by Subject

---

#### Subject No. 1

Subject No. 1 (S1) is a secretary from the Human Services Department. She took the Supervision and Spanish courses during the time of this study. She supervises two clerical people.

Supervision - The course worked for S1. S1 enjoyed the experience and has applied the training to the job. S1 found the instructor excellent and the class lively and informative. The role plays were well received. S1 learned new things, such as there are right and wrong ways to do things as a supervisor. As an example, S1 noted the technique of putting the telephone caller on hold if necessary to maintain control of one's emotions. An example of application of the training to the job was given by S1. It involved dealing with a subordinate who has a chronic attendance problem.

Spanish - The course did not work for S1. S1 took it with a co-worker and both fell behind fast and couldn't keep up. They had trouble competing with several others in the class who had had some Spanish previously. S1 noted, however, the excellence of the instructor. The instructor went out of her way to help S1 and her friend. She even made up special handouts. S1 also said that the handouts were very good. A practice tape was also produced by the instructor. S1 has not used any Spanish on-the-job. S1 refers all Spanish questions or problems to a native Spanish speaking person in the office.

#### Subject No. 2

Subject No. 2 (S2) took the Spanish and Supervision courses during the time period of the study.

Spanish - S2 found the course useful. S2 had prior training in Spanish (in the 7th grade). S2 noted that the material on the Hispanic culture was interesting and well received. S2 would like additional courses in Spanish. S2 can now say more than "Hello," and "One moment please" in Spanish - which S2 said is a positive reflection on the course. S2 now has a definite improvement in her comfort level due to the course. One criticism was that too much time was spent on police related issues because of the large number of police officers in the class. S2 suggested a pre-course survey by the instructor to determine class needs.

Supervision - S2 had a positive reaction overall to Supervision. S2 took the course over one year ago. S2 is a supervisor currently. S2 is a strong believer in training and is the type of student/employee who gets something useful out of every course. No specific job examples were offered by S2.



Subject No. 3

Subject No. 3 (S3) is a Program Director for the Sheriff's Department. S3 took the Intercultural Communications and Supervision courses.

Cultural Relations - This course did not work for S3. S3 expected more discussion and information about the Hispanic Culture that would be relevant to a sheriff's deputy in an arrest situation. S3 found the class to be too general. What was needed, S3 said, was a "how-to class" based on facts about various cultures - Hispanic in particular. No examples given by S3 of use on-the-job. S3 said that most people already knew about the course content, and, if they didn't they were not "reachable" anyway because they are so hardened. S3 had trouble concentrating during class as a result of disappointment with the material. S3 would not recommend course to others. Course was slightly mis-advertised, S3 said, because it promised more than it delivered.

Table 18 (continued)

Supervision - S3 had favorable reaction to the course. S3 said course was delivered true to the advertisement. S3 said role plays were very helpful. Also, the basics were covered well. S3 had no new learning because other courses were taken prior to this. S3 cited an incident when the course helped in dealing with the firing of an employee. S3 said that it also helped increase confidence in dealing with personnel matters.

Subject No. 4

Subject No. 4 (S4) was on the staff of the Animal Shelter. S4 attended Cultural Relations and the Writing courses during the time period of the study.

Cultural Relations - S4 said that the course was not useful, and that, because of background and upbringing, the subject already held many of the accepting views of other cultures that the course was trying to impart. The subject also said that they had expected more of a focus on the predominant characteristics of one or two cultures and the work related impact of those characteristics. S4 said no job duties/tasks have been effected by the course. No comment on class experience per se.

Writing - S4 summed it up as "Excellent class...the most helpful of all I've taken through the city. Excellent instructor, presentation, and course materials. She (instructor) kept the class interesting throughout." Also, S4 said it has been applicable at work because of all the memos and reports they have to write.

Subject No. 5

Subject No. 5 (S5) works as a Mental Health/Mental Retardation Placement Specialist. S5 attended the Writing course during the study time frame.

Writing - S5 had a positive reaction to the course. The teacher was very good, there were several group assignments which made it more interesting. S5 expected more time spent on grammar, however. S5 liked the real world examples of bad and good writing as a learning aid. S5 would recommend course to others. S5 said that the course definitely taught new things, enhanced learning, etc. S5 has applied the course on-the-job mainly in the area of using a better choice of words in writing memos, etc. both business and personal.

Subject No. 6

Subject No. 6 (S6) is in a staff position in the employment section of the Personnel Department. S6 attended the Reading and Writing courses during the study.

Writing - S6 had a very positive reaction to the course. The instructor was just excellent, S6 said. Other strong features were that everyone participated, course length was good, handouts/texts were good, and the exercises were good. Learning took place for S6 because S6 had never taken an English composition course and this was a close approximation. S6 has applied the learning to their job as a writer of job vacancy announcements.

Reading - S6 said that Reading was "OK" but not as good as the Writing course. Instructor was more subdued. It has been more difficult to use the course on-the-job because the course emphasized skimming, and skimming is risky in the type of reading S6 is required to do. No examples of application to the job were given.

Table 18 (continued)

Subject No. 7

Subject No. 7 (S7) is a supervisor of medic personnel in the Fire Department. He attended the Writing course during the study.

Writing - S7 said the course worked, and that the outcome was positive personally. The instructor was cited as being well prepared, dynamic, and competent. S7 could not think of any negatives about the course. Course materials were well received, and exercises and feedback on one's own writing were positive. S7 has really seen a change in personal writing style, has utilized many of the writing tips learned in the course, etc.

**Appendix J**

**Date:** July 29, 1992  
**TO:** All Stakeholders  
**FROM:** John Milatzo

---

Thanks again for meeting with me in May to discuss the study results.

Since our last meeting, I have analyzed all the comments and discussion. I have concluded that the information presented in your notebooks was too much data to be presented and digested at one time, although I did get some very useful information and reactions.

I have decided that a final survey is necessary, however, and this should be the last intrusion on your time -- which all of you have given most generously, and for which I am grateful.

Please complete the enclosed survey within one week if possible. A return envelope is enclosed for your convenience. Please call if you have questions, or if you will be delayed in responding. (202-268-5167, or 703-549-5258). Your continued participation in this study is greatly appreciated!! Thanks again!

John Milatzo

**INSTRUCTIONS:**

Attached is a report of the process we have gone through to date and a description of the results. It is written in the form of an **imaginary memo** from the Personnel Specialist to the Director of Personnel, as if the Personnel Specialist had evaluated five courses and was informing the Director of Personnel of the results.

Please read the memo as if it were distributed to you for written comment. Then answer the questions that follow as best you can, and as you would if you were preparing an actual response to the memo. The Director of Personnel, of course, will answer the questions as the addressee of the memo, whereas the others will answer as interested other parties. Please call if you have any questions.

**DATE:** July 29, 1992  
**TO:** DIRECTOR OF PERSONNEL SERVICES  
**FROM:** PERSONNEL TRAINING SPECIALIST  
**SUBJECT:** EVALUATION OF SELECTED COURSES

---

As you are aware, I have been collecting some evaluation data on some of our courses this year in order to have a greater sense of what impact we are having and where we might improve.

For the past six months, I have been studying five courses: Writing Skills Review (hereafter Writing), Critical Reading (hereafter Reading), Basic Supervision (hereafter Supervision), Intercultural Communications (hereafter Cultural Relations), and Survival Spanish (hereafter Spanish).

My approach involved past users of our training, recent students, their managers, executives and other managers with a stake in the success or failure of the training.

First, I assembled a group of three executives and obtained from them a list of organizational values. I figured that if our training does have impact, it will also be related to organizational values. Conversely, training with less impact ought to be less related to organizational values. The executive panel provided the values and weighted them on importance. Below is a list of the values and their relative importance ratings (expressed as percentages summing to 1.00):

VALUE	WEIGHT
Provide quality service to the citizens; be responsive, maintain a commitment to excellence.	.22
Provide cost-effective service; maintain a balanced budget.	.14
Provide quality administrative support to elected and top city officials.	.19
Practice/promote high ethical standards in both internal and external relations.	.12
Support affirmative action; support EEO and professional development for all employees.	.10
Provide a broad range of support services to the public.	
Provide fiscally conscientious management of city programs and services.	.18

The panel then rated the five courses as to how much their content related to each of the organizational values. A course rated 100 was directly related to a given value, and a course rated zero was unrelated. These scores were then multiplied by the above weights, and those products were summed across values to get a **Strategic Value** score for each course. This score indicates the potential for a course to have organizational impact if the right people attend and if the design and delivery of the training are effective. The Strategic Value scores were:

COURSE	STRATEGIC VALUE
Writing	35.95
Reading	31.34
Supervision	42.31
Cultural Relations	38.39
Spanish	41.38

As you can see, Supervision scored highest on strategic value while Reading scored lowest. Overall, however, there was only a ten point difference between the highest and lowest strategic value score, indicating that there is not a very wide range of difference between the courses on strategic value, although there is some.

The panel then developed and weighted "ideal" course characteristics (called attributes) which, in their judgment as stakeholders and decision makers, represented the criteria by which they would judge whether any of these courses was in fact having an impact on the organization. In other words, the attributes are the characteristics the panel would like to see realized when these courses are delivered to the work force. Note, however, that the attributes may or may not be realized today. However, my investigation was to determine the extent to which they were being realized so that we could take appropriate action. The attributes and their relative importance weights were:

ATTRIBUTES	WEIGHT
Course satisfies important needs	.052
Students learn	.045
High benefit/cost ratio	.042
Instructor effective	.031
Students react positively	.024
Texts/handouts are useful	.021
Learning from other students also takes place	.014
Teaching methods suit content and audience	.010
Positive atmosphere for learning	.007
Job performance improvement results	.232
Aids/tools used on-the-job	.160
Students empowered/confidence enhanced	.125
Job tasks made easier	.107
Professional growth enhanced	.089
Practical application of course outside of work	.035

Job performance improvement back on-the-job was clearly the highest valued attribute by the executive panel, followed by the use of course handouts, job aids/tools, student empowerment, and ease of job task after the course.

Survey questions were then developed to serve as indicators of the attributes. For each attribute, at least one survey question was developed. Each survey question was included in one or more of five surveys: a pre-training student survey, a pre-training survey for the students' supervisors, an end-of-class training feedback survey, and two post-training surveys -- one for students and one for their supervisors.

Surveys were then mailed or administered to students attending these courses from September, 1991 through January, 1992. Two classes of the Supervisor and Spanish courses were included. All other courses held only one class during this period.

All survey responses were averaged and converted to a percentage scale and multiplied by 100 to remove the decimal. These results were then multiplied by the attribute weights to produce a **utility value** for each attribute for each course. When more than one question was used to measure an attribute, all questions were averaged and that average was multiplied by the attribute weight. Finally, utility values were summed across attributes to obtain a **total utility value** for each course.

One final step was used to arrive at a final course impact measure.

A formula was developed to include the following factors: total utility score, strategic value score, and the percentage of employees who, in the minds of a sample of students and their supervisors, have a job-related need for the training. To get a measure of need, incoming students to the above courses and their supervisors were surveyed and asked to give a direct estimate of the percent of employees who have a job-related need for the course in question.

Thus, the survey data allowed for an estimate of organizational impact defined as:  $IMPACT = COURSE\ UTILITY \times STRATEGIC\ VALUE \times NEED$ . The results of these calculations for the five courses, in rank order based on total impact score, were:

TOTAL IMPACT RANK	COURSE	UTILITY SCORE	STRATEGIC VALUE	NEED	TOTAL IMPACT	MAXIMUM IMPACT	% ACTUAL OF MAX IMPACT
1	Cultural Relations	62.89	38.39	.83	1738	2764	63
2	Reading	74.55	31.34	.73	1706	2287	75
3	Writing	75.94	35.95	.61	1666	2193	76
4	Spanish	59.67	41.38	.52	1284	2299	56
5	Supervision	70.10	42.31	.42	1246	1777	70

These results indicate that: a) the total impact for the five courses is spread over a rather broad range, including definite clusters at the high and low ends, b) there is much more variability in (job-related) need than there is in either total utility score or strategic value, c) a course such as Supervision can have a high strategic value and a relatively high utility score, yet have lower impact due to a significantly lower number of people in the work force who have a direct need for it; on the other hand, a course such as Cultural Relations can have only relatively moderate utility and strategic value, but have high impact due to the large numbers of employees it can potentially reach because of the nature of the subject matter. Cultural Relations should be a factor in the work life of most, if not all, employees as these data indicate (need = 83%), d) although the Spanish course has relatively high strategic value, its utility rating and its relatively



low need index place it low on total impact--just slightly above Supervision.

It is interesting, also, to rank order the courses on the basis of total utility score rather than total impact score. This ordering can be viewed as the reverse priority order for needed course improvement since utility score is a direct reflection of the extent to which desired course attributes are being realized. Utility score can also be interpreted as a percentage effectiveness index, since maximum total utility equals 100 in each case. The results were:

UTILITY SCORE RANK	COURSE	UTILITY SCORE	STRATEGIC VALUE	NEED	TOTAL IMPACT
1	Writing	75.94	35.95	.61	1666
2	Reading	74.55	31.34	.73	1706
3	Supervision	70.10	42.31	.42	1246
4	Cultural Relations	62.89	38.39	.83	1738
5	Spanish	59.67	41.38	.52	1284

#### Course Strengths and Weaknesses

The data also allowed for the analysis of individual course strengths and weaknesses. For example, the Writing course results can be studied to see which of the desired course attributes it scored well on, which it scored less well on, etc. leading to recommendations for improvement. Using an arbitrary cut-off of 70 on the average rating from the surveys, attributes were flagged by an asterisk indicating priority areas for improvement. Thus, for the Writing course, the two attributes needing primary attention are 'Job Performance' and 'Practical Application.'

## COURSE: WRITING

ATTRIBUTE	WEIGHT	AVERAGE RATING (0-100)	UTILITY (WGHT. X RATING)
Course satisfies important needs	.052	85	4.48
Students learn	.045	73	3.34
High benefit/cost ratio	.042	92	3.88
Instructor effective	.031	95	3.01
Students react positively	.024	91	2.24
Texts/handouts are useful	.021	89	1.88
Learning from other students also takes place	.014	91	1.28
Teaching methods suit content and audience	.010	93	0.98
Positive atmosphere for learning	.007	93	0.65
Job performance improvement results	.232	58**	13.46
Aids/tools used on-the-job	.160	88	14.14
Students empowered/confidence enhanced	.125	71	8.87
Job tasks made easier	.107	76	8.14
Professional growth enhanced	.089	81	7.23
Practical application of course outside of work	.035	65**	2.32

\*\* Priority for improvement

Attributes needing attention in the remaining courses are listed below:

## COURSE: READING

ATTRIBUTE	WEIGHT	AVERAGE RATING (0-100)	UTILITY (WGHT. X RATING)
Job performance improvement results	.232	53**	12.30
Students empowered/confidence enhanced	.125	70**	8.75

\*\* Priority for improvement

**Comment:** Strategies are needed to improve the extent which students make use of the training on-the-job, and for improving students' confidence in their writing ability.

COURSE: SUPERVISION

ATTRIBUTE	WEIGHT	AVERAGE RATING (0-100)	UTILITY (WGHT. X RATING)
Students learn	.045	64**	2.92
Job performance improvement results	.232	50**	11.60
Students empowered/confidence enhanced	.125	61**	7.62
Practical application of course outside of work	.035	61**	2.17

\*\* Priority for improvement

**Comment:** Strategies are needed to improve the extent to which students learn the material, make use of the training on-the-job, and for improving student confidence in their supervisory functions and applying the training in their personal life.

COURSE: CULTURAL RELATIONS

ATTRIBUTE	WEIGHT	AVERAGE RATING (0-100)	UTILITY (WGHT. X RATING)
Students learn	.045	63**	2.88
Job performance improvement results	.232	48**	11.14
Students empowered/confidence enhanced	.125	55**	6.87
Job tasks made easier	.107	58**	6.21
Professional growth enhanced	.089	56**	5.0
Practical application of course outside of work	.035	60**	2.14

\*\* Priority for improvement

**Comment:** Strategies are needed to improve the extent to which students learn the material, make use of the training on-the-job, gain confidence and empowerment, conduct job tasks with greater ease, realize professional growth, and apply the training in their personal life. Realistically, this type of training is very hard to translate into performance. It is mainly awareness training.

## COURSE: SPANISH

ATTRIBUTE	WEIGHT	AVERAGE RATING (0-100)	UTILITY (WGHT. X RATING)
Job performance improvement results	.232	44**	10.21
Students empowered/confidence enhanced	.125	52**	6.5
Job tasks made easier	.107	47**	5.03
Professional growth enhanced	.089	54**	4.82
Practical application of course outside of work	.035	41**	1.46

\*\* Priority for improvement

**Comment:** Strategies are needed to improve the extent to which students make use of the training on-the-job, gain confidence and empowerment, conduct job tasks with greater ease, realize professional growth, and apply the training in their personal life. In contrast to Cultural Relations, Spanish appears to be in greater need of examination to determine whether the right population is being trained, and whether the level of material is appropriate given the objective of enabling students to interact better with Spanish speaking populations.

In summary, I have drawn the following conclusions from the data: 1) that the five courses do vary on their strategic value potential for the organization, but the range of difference is small (approx. ten points); 2) there was some notable variation in the extent to which there is a perceived need for these five courses (83% for 42% for Supervision); 3) application of the training back on-the-job and related attributes is clearly the most highly valued training outcome for the executive panel; 4) Cultural Relations is having the highest overall course impact, Spanish and Supervision the lowest, with Writing in between; 5) when only the extent to which the courses are realizing the desired attributes is considered (i.e. no regard for percent need or strategic value), Reading and Writing have the greatest utility with Cultural Relations next to the lowest; 6) except for two cases, course outcomes (i.e. things that happen back on-the-job after training) are clearly where most of the effort should be placed to improve overall impact.

I offer the following recommendations for your consideration as we approach FY 93:

**RECOMMENDATIONS:**

1) Explore/develop strategies for realizing a greater number of desired attributes for the Cultural Relations, Spanish, and Supervision courses. These three courses show the lowest total utility score which is a direct reflection of attribute deficiencies.

Focus will be on developing ideas and action items for increasing job performance and related attributes pertaining to the job application of training.

- 2) Examine the Writing course jointly with the instructor and explore ways to effect greater application of the course material on-the-job and in the personal life of the students.
- 3) Examine the Reading course for opportunities to effect greater job performance application as well as increasing empowerment and student confidence.
- 4) Explore the question of student learning in the Supervision course. The relatively low learning score could mean weak content in terms of depth/breadth. Or, it may reflect the basic nature of the material and the fact that the mandatory nature of the course attracts too many experienced supervisors for whom the content is redundant. Perhaps the population attending may need tighter screening for need.
- 5) The "awareness" aspect of the Cultural Relations course could mean that not much can be done to increase the use of impact of the training on outcome attributes such as job performance, tasks made easier, etc. However, discussions with the instructor on this issue is recommended.
- 6) Spanish would appear, on the surface, to be a strong candidate for high application and transfer to the job. I recommend a thorough analysis of who is attending this training and discussions with former students and the instructor to focus on barriers to application to the job.

I would like your comments and recommendations in response to this memorandum before I initiate further action.

Personnel Training Specialist (see follow-up questions attached)



2. Study Results --

a) Were the study results credible? For example, given the process that was used and the results that were found, do you have any difficulty believing that the Cultural Relations course, for example, is having the greatest impact (1738) or that it is only having 63% of the impact that it could have?

---

---

---

---

---

---

---

b) How do you think the highest levels of management would view the credibility of the results?

---

---

---

---

---

---

---

c) What, if anything, is missing from the memo/report that should have been included?

---

---

---

---

---

---

---

3. Process --

a) Did the effort to obtain impact from various stakeholders seem appropriate, overkill, underkill, etc.?

---

---

---

---

---

b) Did the development of the strategic value score, course attributes and weights, final utilities, and overall impact calculations seem: appropriate, too complicated, too simple, scientific, unscientific, etc.?

---

---

c) Having participated in this process as stakeholders, what were the strengths of the process? What were the weaknesses?

---

---

---

---

---

d) Is this the way courses should be evaluated in this organization?

---

---

---

---

---



4. Utilization --

a) What would happen, typically, as a result of this memo? Would further investigation be authorized? Why, why not?

---

---

---

---

---

---

b) Would the information be shared with the highest levels of management? Why, why not?

---

---

---

---

---

---

c) Would/could the information be used successfully to defend the impact of these courses in the organization? Why, why not?

---

---

---

---

---

---



**PART B**

Attached are suggested "desired" attributes (or standards for) of a course evaluation process.

Step 1 -- Assign an importance rank to each attribute as it reflects importance to you as a stakeholder. (1 = most important; 4 = least important).

Step 2 -- Assign the attribute ranked 4th the weight of 10. Then assign the attribute ranked 3rd a weight of 10 or higher expressing how much more important it is than the first attribute. Move up the ranks weighting each attribute in relation to the first (e.g. "20" = twice as important as "10")

Step 3 -- Rate the process used to evaluate the five courses on each of the four attributes. Under the column marked "Rating," assign a rating of the extent to which the process satisfied the desired attribute. Rate the process on each of the attributes. "100" equals maximum, "0" equals minimum.

RANK	DESIRED EVALUATION ATTRIBUTES (i.e. STANDARDS)	WEIGHT	RATING
	<p><b>UTILITY:</b> Evaluation will serve the practical information needs of given audiences. Utility standards include evaluator credibility, focus on pertinent questions, clear description of methods used, report timeliness, and conducted in ways that encourage follow-through by relevant audiences.</p>		
	<p><b>FEASIBILITY:</b> Evaluation will be realistic, prudent, politically sensitive, and cost effective.</p>		
	<p><b>PROPRIETY:</b> Evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved as well as those effected by the results.</p>		
	<p><b>ACCURACY:</b> Evaluation will reveal and convey technically adequate information about features of the objects (i.e. courses) being evaluated that determine their merit and worth.</p>		

## Appendix K

### PART A - Follow-up Questions to Imaginary memo.

Please answer each question in the space available.

1. What would your comments/recommendations be if this were a real memo and comments were required?

#### Respondent #1:

My comments on each of the five recommendations would be as follows:

- 1) I think you have "the cart before the horse!" If a course has low utility, replace it with one with higher utility.
- 2) Excellent suggestion, although I would de-emphasize the "personal life of the students" element.
- 3) Good idea; go ahead.
- 4) Good idea; go ahead.
- 5) Good idea; go ahead.
- 6) Good idea; go ahead.

#### Respondent #2:

Bill's recommendations:

- 1) I agree. These courses are important and need to be included in the Training Plan on an ongoing basis.
- 2) At this point in time, it would be better not to address personal goals.
- 3) Good idea.
- 4) Good idea.
- 5) Good idea.
- 6) Good idea.

#### Respondent #3:

You reserved a great deal of space for a very short answer. My response would be "go for it" reference the recommendations. It seems like the obvious thing to do. Since that much work went into the study and specific conclusions have been reached, the only payoff is to follow up on the recommendations.

If I were the Personnel Director, I would ask how we will determine that the changes, once made, had the impact they were supposed to have--short of starting completely over from square one. Maybe square two since the values assigned by the stakeholder groups would remain the same. In other situations, the Personnel Director would never ask for this much detail. The recommendations would be accepted with reservations if Peter was not familiar with the process.

Reading this memo is like sleeping with a lawyer (huh?). Too much jargon. I have this tune stuck in my head but I can't remember

the words. Do I know the song or not?

Respondent #4:

Overall, this is very helpful in evaluating and planning for training. More operational definition at key concepts/terms is needed; glossary and/or more explanation in body of memo - it is easy to get them confused.

- o The course strengths and weaknesses section is most practical and clear.
- o An early overview of your "hypothesis" or approach to what follows would help out reader. Otherwise, you get caught up in math steps without proper grounding.

2. Study Results --

a) Were the study results credible? For example, given the process that was used and the results that were found, do you have any difficulty believing that the Cultural Relations course, for example, is having the greatest impact (1738) or that it is only having 63% of the impact that it could have?

Respondent #1:

I think the credibility of the results are clouded by the complexity of the evaluation process. I have no factual basis for challenging the results, however.

Respondent #2:

Because so much time elapses between discussions on the process, it is difficult to understand. I do not have any reasons to believe the results are not credible.

Respondent #3:

If I understand this, which I doubt, it would be the 63%. I don't understand what the total impact score indicates and I've read this several times. Following our conversation--I have no problem believing any of the results now that I have a better understanding of the whole process.

Respondent #4:

Yes - assuming you are comfortable with your multi-step calculations early in the memo.

b) How do you think the highest levels of management would view the credibility of the results?

Respondent #1:

With mild interest that would not be sufficient to override their sense (intuitive) of the utility of each individual course.

Respondent #2:

The credibility of the results would not be a factor in their deciding which classes should be offered.

Respondent #3:

If it were given to them in this detail, they would lose patience and toss it. Wrapped differently, I think the results would be accepted but the resulting recommendations not necessarily approved. (See 4b)

Respondent #4:

The methodology could be problematic, unless it is simplified/"packaged" in a more understandable way.

c) What, if anything, is missing from the memo/report that should have been included?

Respondent #1:

A discussion of how we can institutionalize the evaluation technology: expand it to all courses, implement with current resources/staff, and simplify without compromising quality.

Respondent #2:

The report is too complex and needs to be simplified. It should spell out how we intend to evaluate future courses on an ongoing basis.

Respondent #3:

In an actual memo, the reason these particular courses were chosen for study would be stated or restated. That's not important for this exercise. Also, the way courses have been evaluated in the past (or have not been). A one paragraph explanation of the process with a full explanation as an attachment.

Respondent #4:

As mentioned: 1) initial conceptual "roadmap"; 2) glossary and operational definitions; 3) more examples of calculations as each step is introduced.

## 3. Process --

a) Did the effort to obtain impact from various stakeholders seem appropriate, overkill, underkill, etc.?

Respondent #1:

A good cross-section of staff; complexity of survey instruments at times seemed like overkill.

Respondent #2:

The selection of stakeholders was good and appropriate.

Respondent #3:

Certainly not overkill. In order to obtain a reliable answer, you have to look at the issue from all sides. Since all of this is a matter of perception, it seems appropriate to quantify the perceptions of a group of subject matter experts.

Respondent #4:

Appropriate - this was the "rubber hits the road" and was the highest rank attribute.

b) Did the development of the strategic value score, course attributes and weights, final utilities, and overall impact calculations seem: appropriate, too complicated, too simple, scientific, unscientific, etc.?

Respondent #1:

Too complicated.

Respondent #2:

Too complicated.

Respondent #3:

Have to say too complicated - everything is a blur at this point. The development of everything was too time consuming. I think it is a bit too complicated but at the same time, appropriate.

Respondent #4:

Too complicated as presented.

c) Having participated in this process as stakeholders, what were the strengths of the process? What were the weaknesses?

Respondent #1:

**Strengths** - group discussion/participation

**Weakness** - complexity; time required to fully understand the evaluation process



Respondent #2:

**Strengths** - input from participants

**Weaknesses** - time consuming and complicated

Respondent #3:

**Strengths** -Covers all sides and seems to be very thorough. Reducing abstract information to numbers should make the results more easy to analyze.

**Weaknesses**-Info overload perhaps. Got lost on the way to Grandma's house and the wolf ate me. Also time consuming.

Respondent #4:

**Strengths** -interesting, challenging, your flexibility

**Weaknesses**-not enough time; never felt fully grounded in concepts

d) Is this the way courses should be evaluated in this organization?

Respondent #1:

If there is a way to simplify the process.

Respondent #2:

It would be worthwhile discussing. It does need to be simplified.

Respondent #3:

Yes and no. This is the best way to gather and quantify the data. More reliable than guesstimation or taking a best shot. On the other hand (which has diamond rings and nail polish) what the f\_\_\_!

Respondent #4:

Some of your methods for sure. Not certain about full methodology.

4. Utilization --

a) What would happen, typically, as a result of this memo? Would further investigation be authorized? Why, why not?

Respondent #1:

I would authorize the Personnel Training Specialist to take the steps he suggests in "Recommendations." (See #1)

Respondent #2:

The Personnel Director would give his approval and Bill and I would plan the process.

Respondent #3:

The Personnel Director would authorize further investigation time permitting. I'm not sure what type of further investigation would be appropriate or warranted. The memo ends with a request for comments before the initiation of further action but does not state what it would be. A decision maker would have to know that.

Respondent #4:

Me and the Personnel Director, and perhaps Training Committee, would meet and discuss; eventual submission to vola.

b) Would the information be shared with the highest levels of management? Why, why not?

Respondent #1:

I would share the information only in the context of presenting a new FY training plan, and only if there was disagreement over the courses we should offer.

Respondent #2:

The Personnel Director would make this decision. I would recommend that we share the information.

Respondent #3:

The total analysis would not be shared. It would confuse and confound. The bottom line would be shared and questions would be answered regarding the process. The highest levels of management are involved in training decisions and would be informed.

Respondent #4:

Yes, after Peter digests, adapts/revises.

c) Would/could the information be used successfully to defend the impact of these courses in the organization? Why, why not?

Respondent #1:

Yes, it could be used, but we would need a way to describe the technology simply and quickly before people lose interest.

Respondent #2:

Yes, it could be used. However, we would need to be able to explain it very simply.

Respondent #3:

We could try. The results are certainly more tangible than "gut feelings" and "shooting from the hip" techniques of defense used in the past. But this would require explaining the process which is cumbersome and few would sit still for that. I see problems here.

Respondent #4:

Could be an important factor, along with less objective factors (Manager's perceptions/feelings)

d) Other: Please provide any other comments relevant to the use/non-use of the information in the memo, or to the strengths/weaknesses of the process used to obtain the information.

Respondent #1:

Process used to obtain information was fine; it would be interesting to check reliability among participants: 1) Do participants ranking of weights/factors change over time? 2) Do variances among rates show a pattern?

\* I hope that we will be taken seriously on the request to simplify the evaluation technology so that we can institutionalize some form of improved training evaluation process.

Respondent #2:

We have needed a training evaluation process for a long time. I believe what you have developed is very worthwhile and could and should be used for all courses offered by the organization. We would need to shorten the process in some way and make it easier to understand.

Respondent #3:

As we mentioned on the phone, the scores take us in a circle. The utility score is as far as I would need to go. The % of max impact returns to the utility score - so why go through the other steps to get there?

We also discussed "par." I like this procedure because it reduces abstract information to something you can get a handle on - numbers. But at the end, you must still guess the point at which improvement is needed. On my old evaluation form a participant scores the class on a scale of 5=Excellent to 1=Poor. Since I want all of my classes to be better than average, anything over a 3 accomplishes that.

My target, however, is Very Good (4). If I get an average of less than 4, I read the comments and deduce the needed changes. How can I relate what was said or felt about a class to the utility score? How do they relate?

Respondent #4:

You have done a very professional job on this, Peter. The only real problem was our fragmented time with you. Getting a commitment from us for a whole day; 2 half days; etc. would have worked better - but is very hard to do. My conceptual gaps are due to this fragmentation primarily.

PART B

Attached are suggested "desired" attributes (or standards for) of a course evaluation process.

Step 1 --Assign an importance rank to each attribute as it reflects importance to you as a stakeholder. (1 = most important; 4 = least important).

Step 2 --Assign the attribute ranked 4th the weight of 10. Then assign the attribute ranked 3rd a weight of 10 or higher expressing how much more important it is than the first attribute. Move up the ranks weighting each attribute in relation to the first (e.g. "20" = twice as important as "10")

Step 3 --Rate the process used to evaluate the five courses on each of the four attributes. Under the column marked "Rating," assign a rating of the extent to which the process satisfied the desired attribute. Rate the process on each of the attributes. "100" equals maximum, "0" equals minimum.

**Respondent #1:**

RANK	DESIRED EVALUATION ATTRIBUTES (i.e. STANDARDS)	WEIGHT	RATING
3	<b>UTILITY:</b> Evaluation will serve the practical information needs of given audiences. Utility standards include evaluator credibility, focus on pertinent questions, clear description of methods used, report timeliness, and conducted in ways that encourage follow-through by relevant audiences.	15	55
2	<b>FEASIBILITY:</b> Evaluation will be realistic, prudent, politically sensitive, and cost effective.	20	40
4	<b>PROPRIETY:</b> Evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved as well as those effected by the results.	10	80
1	<b>ACCURACY:</b> Evaluation will reveal and convey technically adequate information about features of the objects (i.e. courses) being evaluated that determine their merit and worth.	30	50

**Respondent #2:**

<b>RANK</b>	<b>DESIRED EVALUATION ATTRIBUTES (i.e. STANDARDS)</b>	<b>WEIGHT</b>	<b>RATING</b>
3	<b>UTILITY:</b> Evaluation will serve the practical information needs of given audiences. Utility standards include evaluator credibility, focus on pertinent questions, clear description of methods used, report timeliness, and conducted in ways that encourage follow-through by relevant audiences.	20	50
1	<b>FEASIBILITY:</b> Evaluation will be realistic, prudent, politically sensitive, and cost effective.	30	50
4	<b>PROPRIETY:</b> Evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved as well as those effected by the results.	10	90
2	<b>ACCURACY:</b> Evaluation will reveal and convey technically adequate information about features of the objects (i.e. courses) being evaluated that determine their merit and worth.	25	60

**Respondent #3:**

RANK	DESIRED EVALUATION ATTRIBUTES (i.e. STANDARDS)	WEIGHT	RATING
2	<b>UTILITY:</b> Evaluation will serve the practical information needs of given audiences. Utility standards include evaluator credibility, focus on pertinent questions, clear description of methods used, report timeliness, and conducted in ways that encourage follow-through by relevant audiences.	15	60
4	<b>FEASIBILITY:</b> Evaluation will be realistic, prudent, politically sensitive, and cost effective.	10	5
3	<b>PROPRIETY:</b> Evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved as well as those effected by the results.	12	80
1	<b>ACCURACY:</b> Evaluation will reveal and convey technically adequate information about features of the objects (i.e. courses) being evaluated that determine their merit and worth.	20	70

**Respondent #4:**

RANK	DESIRED EVALUATION ATTRIBUTES (i.e. STANDARDS)	WEIGHT	RATING
1	<b>UTILITY:</b> Evaluation will serve the practical information needs of given audiences. Utility standards include evaluator credibility, focus on pertinent questions, clear description of methods used, report timeliness, and conducted in ways that encourage follow-through by relevant audiences.	40	80
2	<b>FEASIBILITY:</b> Evaluation will be realistic, prudent, politically sensitive, and cost effective.	30	50
4	<b>PROPRIETY:</b> Evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved as well as those effected by the results.	10	90
3	<b>ACCURACY:</b> Evaluation will reveal and convey technically adequate information about features of the objects (i.e. courses) being evaluated that determine their merit and worth.	20	75



**A User's Guide to  
Multi-Attribute Utility Technology (MAUT)**

CONTENTS

- I. Introduction to MAUT
- II. The MAUT Steps
- III. Preparing for a MAUT Evaluation
- IV. Conducting the MAUT Workshop
- V. Analyzing the MAUT Data
- VI. Reporting the MAUT Results

## I. Introduction to MAUT

Multi-Attribute Utility Technology (MAUT) first appeared in the program evaluation literature in 1975 as an alternative to the experimental method for program evaluation.

The main foci of MAUT are the decisions to be made and the values of those who will make those decisions. Values are relevant in MAUT because values are inextricably linked to the criteria decision-makers use to evaluate programs, and evaluation criteria (also called dimensions or attributes in MAUT) are the means and basis by which the decision maker(s), called stakeholders, make their decisions about programs.

Subjective as decision criteria may be, the MAUT process captures, quantifies, and makes explicit the decision criteria that are inextricably part of the judging, evaluating, and decision making process. As such, MAUT has roots in decision theory and Bayesian statistics. Bayesian statistics are based on a theory of conditional probability which differs from classical statistical theory. Baye's theorem is a means of combining new information with prior information to produce revised probability estimates of events. The main difference between classical and Bayesian statistics is that Bayesians are interested in the inverse of the problem of deducing the consequences of specified hypotheses. In short, classical statisticians are concerned with deducing conclusions from hypotheses, whereas Bayesians are concerned with drawing conclusions about hypotheses from observations of consequences.

MAUT can incorporate any information developed from experimental or quasi-experimental research, if it is available. Classical research results are accepted and can be accommodated in a MAUT decision making system, but much consideration is given to the human situation that goes beyond classical experimental methods, and allows the use of data that are relevant and specific to a particular decision, even though they may be too "subjective" for most advocates of the more classical approaches.

From decision theory comes the notion that there are four main phases to the decision process. They are: (1) recognition of the decision problem; (2) probability evaluation; (3) evaluation--attachment of values to outcomes; and, (4) choice among acts. The influence of these elements can be seen clearly in the steps of the MAUT process outlined below.

The other influence on MAUT comes from the assumptions and principles underlying Bayesian statistics. In the Bayesian/decision theoretic model, the role of the evaluator is that of a facilitator for decision making. The evaluator collects data and presents it to the decision maker who will make a decision. The most important question for the decision maker is "What data should be collected?" The Bayesian-decision theoretic model is most useful for answering this question because it is derived from the assumption that people make decisions by evaluating various entities (alternatives) on many relevant value dimensions.

If two programs need to be compared, for example, program effectiveness will be an important dimension. But in the typical classical/experimental design, effectiveness will likely be the only criterion. In the contrasting Bayesian model, many additional important factors can be accommodated. The model acknowledges the multi-faceted complexity of decision making. It attempts to quantify the decision process by isolating the values held by the decision maker and prioritizes them in the way he or she does when making a decision. Data are then collected on the important dimensions, and the data are used to modify prior expectations into a newly integrated assessment of facts and probabilities about states of the world. Thus, a decision is made using the best information available which, in the real world, is almost always a combination of old and new information.

The relevance of the Bayesian concept for the application of MAUT to evaluation and decision-making is that MAUT can be easily used in multiple instances over time. When new data are acquired, MAUT utilities can be recalculated to produce new utilities which merge new data with old data in a manner analogous to how people make rational decisions in daily life.

Unlike the classical statistics approach, no set of data can (or should) stand alone in a Bayesian system--especially with rare or small effects. Some writers have said that a new model was needed to replace the classical statistics model, one that allowed data collection to be continuous, intermittent, and/or irregular as illustrated in the above example of merging new data over time. They see Bayesian statistics as providing the link for the merging process.

## II. The MAUT Steps

The essentials of the MAUT method are usually described as series of developmental and analytical steps:

1. Identify the objects of evaluation (called entities), or the functions or processes which the evaluation is to perform.
2. Identify the stakeholders--Those people who have an interest or "stake" in what is being evaluated and/or who are important enough so that their interests should be considered.
3. Elicit from stakeholders a common set of the relevant value dimensions or attributes which the program(s) being evaluated must/should hold. These are the "criteria" by which the various stakeholder groups do and will judge the program(s).
4. Elicit from each stakeholder group the relative importance (i.e., weights) of the items comprising the set of attributes or value criteria.
5. Identify, through empirical research, or systematic subjective ratings, how well each program serves each value

or attribute criteria. (called establishing location measures)

6. Aggregate the location measures with measures of importance. Essentially, this involves multiplying each location measure times the weight of the criterion or attribute and summing across attributes to get a utility value for each program.
7. Perform a sensitivity analysis by manipulating the numbers to see what effect increasing the value of any attribute or increasing the location measure will have on the utility index.

MAUT is viewed as well suited to meet the needs of training evaluation. Advocates of MAUT claim that its advantages for training evaluation are many. Among the advantages are: (1) the decision-based focus, (2) the inclusion and recognition of decision maker values, (3) the quantitative structure that allows for comparisons among alternatives, (4) the ability to handle changing values, (5) the ability to handle experimental data, (6) the ability to merge new data with old through Baye's theorem, and (7) the ability to produce structured evaluative information quickly.

### III. Preparing for a MAUT Evaluation

It is tempting to think of a MAUT evaluation as beginning when a group of stakeholders are assembled. But there are important things to consider prior to that.

**Identifying the Client:** The first is to make sure the client for the evaluation is clearly identified. It is important to include the client for the study as much as possible in the MAUT process. This will maximize the likelihood that the results will be used to influence decision making. To the extent possible, make sure the evaluation is really wanted. The evaluation may be doomed to fail if the evaluator is the only client.

**Identifying the Decision Maker(s):** It is also important to find out who the real decision maker for the program is. This person(s) may or may not be the client for the study. If the real decision makers(s) are not included, the study will lack a critical component--namely, values of the person(s) whose values need to be maximized by the program improvement efforts which hopefully will follow the evaluation.

If the primary decision maker(s) cannot participate as stakeholders in the group process, try to interview them and get their values and expectations about the program.

**Identifying the Stakeholders:** Another important issue is who are the stakeholders and how are they identified. In general, stakeholders are people who have an interest and stake in the program being evaluated, and/or a stake in the outcome of the evaluation.

Stakeholders can come from a variety of places. The questions below will help you identify stakeholders. The number and types you ultimately include in the study, and how you group them will depend on time and resources available to you (and to them). The ideal would be to have several stakeholder groups, each rather homogeneous in makeup so that each group is representative of a particular stakeholder population. However, there are also advantages to having a group of stakeholders reflecting a variety of sub-populations and points of view. This will provide opportunities for value conflict and clarification to take place about the program. Communication can be enhanced between principal parties and players surrounding the program.

To help in identifying stakeholders, some relevant questions in the training evaluation context are: 1) Who determines what courses are taught?, 2) Who determines course content?, 3) Who approves funds/budget for the training?, 4) Who directs the training staff?, 5) Who does the instruction?, 6) Who attends the training?, 7) Who manages/supervises the trainees?, 8) Who receives the products produced by the trainees?

Thus, policy decision makers, program managers, participants, supervisors of participants, and clients of participants all have a role or stake in the training, all in somewhat different ways. The challenge for the evaluator is to bring as many stakeholders or stakeholder interests into the evaluation as possible.

**Stakeholder Ability to Perform the MAUT Tasks:** Evaluators should be aware that program stakeholders will vary in their ability to perform the necessary tasks of the MAUT process. One solution to this would be to aim for the highest level stakeholders possible in the organization for at least one of the stakeholder groups. If only one stakeholder group is possible, it is suggested that the evaluator aim for the highest levels possible for at least some of the group members. This of course needs to be balanced against the need for other stakeholder sub-populations to have a voice and input to the evaluation.

Thus, the evaluator needs to be continually aware that although broad representation of stakeholders is desirable, valuable time will be wasted to the extent that evaluators cannot deal effectively with such tasks as developing and weighting desired program attributes. Perhaps the evaluator can control this somewhat by interviewing prospective stakeholders individually prior to selection.

#### IV. Conducting the MAUT Workshop

The role of the stakeholders is critical to the evaluation. When possible, it is recommended that at least one group of stakeholders participate as evaluation team members under the direction of the evaluator. This will help insure full participation and commitment of stakeholder time and attention or attitude.

The MAUT workshop is the name given to the session(s) wherein stakeholders conduct the necessary activities of MAUT, i.e. attribute development, ranking, weighting, indicator development, prior location measures, etc.

There are several ways to start stakeholders thinking about desired program attributes. One way is to ask them: "When all the data are in, what would you like to be able to say about this program?" Another way is to use a fill-in-the-blank question: "The problem with training around here is that \_\_\_\_\_.", or, "When I think about this course, what concerns me most is \_\_\_\_\_".

General brainstorming about desired course attributes is another way to start the attribute development process.

Another starting point is the stated objectives of the program. Formal objectives can certainly be used to start the process, however, stopping at the program objectives is very risky. The main reason is that you may fail to capture the values of the particular stakeholders identified for the evaluation you are conducting. Formal program objectives often reflect only the values of a small and select number of stakeholders, usually program developers/administrators.

**Scheduling the MAUT Workshop Sessions:** It is recommended that evaluators plan for two four-hour sessions with each stakeholder group. Experience has shown that more than four hours is needed to cover the essential of the MAUT process. That process should include attribute development, ranking and weighting of attributes, ranking and weighting of attribute branches, development of attribute indicators, ranking and weighting of attribute indicators (optional), and development of prior location measures.

It is recommended that the first four hour session include an introduction to MAUT, attribute development, ranking and weighting. The second session can include indicator development and prior location measures. Session number two can also include the specification of any standards of acceptability for utility values, general or specific. Research has shown that user's of the MAUT information prefer a priori standards to help them interpret MAUT utility values. These can be provided by simply asking stakeholders, prior to empirical data collection, to specify what percent of the maximum possible utility value for each attribute, or for the overall utility score, would be acceptable in order for the program to be judged "worthwhile".

**Collecting Empirical Data in MAUT:** After the first two MAUT sessions have been conducted, and if there is interest in collecting evaluative data from other sources (either because the clients for the study insist, or because the attribute indicators call for such), it is recommended that the focus group procedure be employed over surveys in the interest of time. For example, former students of a course under evaluation could be assembled quickly and asked to rate the course on the various attributes as a way to obtain post location measures.

Post location measures, it will be recalled, measure the extent to which the program is currently realizing the desired attributes. Much time can be saved in getting relevant information in the focus group setting as compared with time consuming surveys. With the use of a laptop computer and a Lotus program, the evaluator can literally collect post location data in a focus group setting, key it into the software, and produce utility values for decision making on the spot, thus realizing one of the stated advantages of MAUT, i.e., timely evaluation information.

#### V. Analyzing the MAUT Data

**Calculation of MAUT Utilities:** The calculation of utilities a straight forward process. The arithmetic is simple, although it can look more complicated than it is. Essentially, each attribute will eventually have a post- location measure which, when multiplied by the standardized attribute weight, will produce an attribute utility value. When each attribute utility value is summed, the result is the general or overall utility value. The process is a little more complicated in the case where attributes are grouped into branches.

Table 1 presents a sample set of attributes, weights, location measures, and utility values for a hypothetical evaluation of a training course. The attributes for this course are grouped into two branches, a Process Branch and an Outcomes Branch. The process branch is weighted .25, the Outcomes branch is weighted .75. Stakeholders originally weighted the Process branch 10, and the Outcomes branch 30 on importance. Ten divided by 40, the sum of the weights, gives .25, etc.

Column (a) gives the raw importance weights for each of the attributes within each branch. The least important attribute is rated 10, the next least a rating relative to the ten such that it indicates how much more important than the first attribute is the second, and so on. Column (b) gives the within branch normalized weights by dividing each weight in column (a) by the sum of the weights in column (a) within each branch. Column (c) gives the overall standard, or total weights by multiplying each weight in column (b) by the branch weight of its respective branch.

Column (d) gives the location measure of each attribute (scale = 0 to 100). These could be either prior or post location measures. Column (e) gives individual and total utility values arrived at by multiplying the values in column (d) by their respective weight from column (c).

Thus, in this example, the utility of this course, based on the attributes shown and location measures arrived at through stakeholder input and concurrence, is 75.9 which represents approx. 75% of maximum possible utility value (assuming, of course, that 100% on each location measure is actually possible!).



Table 1  
 Decision Value Tree - Executive Stakeholder Group  
 Sample Data from a Single Training Course

Branch/Twig	Weight (a)	Stand- ardized Weight (b)	Total Weight (c)	Location Measure (d)	Utility (e)
Process (Weight = .25)					
Content aims at important needs	75	.21	.052	85	4.4
Learning is insured/knowledge or skill is gained	65	.18	.042	73	3.3
High benefit/cost ratio	60	.17	.042	92	3.8
Effective, competent, knowledgeable instructor	45	.13	.032	95	3.0
Well received/enjoyed by students	35	.10	.025	91	2.2
Useful text, handouts, etc.	30	.08	.020	89	1.8
Learning from other students possible	20	.06	.015	91	1.2
Teaching method matches material and audience needs	15	.04	.010	93	0.9
Lively, energetic atmosphere	<u>10</u>	.03	.010	93	0.6
	355				
Outcomes (Weight = .75)					
Increases current job performance	65	.31	.232	58	13.4
Provides useful job-aids/tools	45	.21	.157	88	14.1
Students empowered w/heightedened self-esteem, confidence, competence, etc.	35	.17	.127	71	8.8
Job tasks made easier	30	.14	.105	76	8.1
Contributes to professional growth	25	.12	.090	81	7.2
Practical application beyond the job	<u>10</u>	.05	.037	65	2.3
	210				75.9

(a) = Attribute Weight (c) = Standard Weight X Branch Weight (e) = c X d

(b) = Weight/Branch Total (d) = Location Measure (100=HI, 0=LO)

**Use of Lotus Software:** Those familiar with the capabilities of Lotus 1-2-3 software will recognize the potential for a Lotus spreadsheet to aid in the calculation of MAUT utilities. The data in Table 1 was produced originally using a Lotus spreadsheet. Lotus is highly recommended by the writer for use in MAUT analysis, particularly because of its ability to assist in sensitivity analyses.

**Sensitivity Analysis:** Sensitivity analysis is a process that MAUT advocates recommend that evaluators go through with the clients of the evaluation. It consists of inserting various alternative weights and location measures into the MAUT decision table to see the extent to which the utility values change with changing conditions of attribute weight and attribute location measure. The process is useful in both decision making and evaluation to aid in the development of program improvement strategies. With Lotus, one can see immediately the changes to utility values that come about by making various changes to attribute weights. This is also useful in feeding MAUT results back to stakeholders in a focus group or briefing format.

#### VI. Reporting the MAUT Results

A quality MAUT evaluation by any other standard will be useless to no one if the right information is not communicated to the right people, at the right time, and in an understandable manner.

Traditional technical reports or even executive summaries can be turn-offs to decision makers if they appear too technical or foreign to the reader who wants to act on the results. Even the simplest of executive summaries in the mind of the evaluator may look exceedingly complex to the user of the information.

It is therefore recommended here that at the very least a two-pronged approach to reporting results be considered. The primary method would be a face to face briefing by the evaluator or evaluation team. The secondary would be a follow-up memo with highlights from the briefing material.

In the briefing, only two or three tables should be used to talk from and illustrate points. Table 1 above is an example of a useful table and, together with the associated table description, provides an example of how to present MAUT information. Another obvious advantage of the briefing format is that attendees can ask questions and get clarification not possible with just the written report.

Finally, the follow-up written report can take the form of an executive summary now that the briefing has taken place. The final report should be as brief as possible, and only include charts and tables that have been covered in the face to face setting.

In addition, it should be noted that a unique and valuable benefit of MAUT is that it brings the evaluator in close contact with the social and political culture that surrounds the program under

evaluation. The report should therefore include critical observations by the evaluator or evaluation team which describe key findings of the social and political nature that will help to make program improvement strategies as effective as possible.

Biographical Vita

EDUCATION: Diploma, Beloit Catholic High School, 1963; B. S. (Psychology), University of Wisconsin (Madison), 1967; M. A. (Industrial Psychology), George Washington University, 1974; Ph.D. (Educational Research and Evaluation), Virginia Polytechnic and State University, 1993.

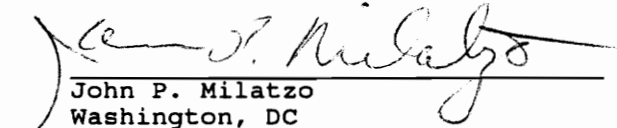
EMPLOYMENT: 1967-1977, National Security Agency, Ft. Meade Md.  
Positions:

Psychology Technician  
Research Psychologist

1977-Present, U.S. Postal Service (Headquarters),  
Washington, DC

Positions:

Psychologist  
Principal Psychologist  
Program Manager: Educational Research and  
Evaluation  
Manager, Instructional Design Branch  
Principal Math Statistician  
Program Manager, Data Collection Management



John P. Milatzo  
Washington, DC  
May, 1993