DMDC TECHNICAL REPORT 95-004

COST-EFFECTIVENESS ANALYSIS OF TRAINING IN THE DEPARTMENT OF DEFENSE

Henry Simpson

JUNE 1995

DMDC

DEFENSE MANPOWER DATA CENTER
Training & Readiness Evaluation and Analysis Division

This report was prepared for the Office of the Deputy Under Secretary of Defense for Readiness by the Defense Manpower Data Center, Seaside, CA. The views, opinions, and findings contained in this report are those of the author and should not be construed as an official Department of Defense position, policy, or decision unless so designated by other official documentation.

DEFENSE MANPOWER DATA CENTER
DoD Center Monterey Bay • 400 Gigling Road • Seaside, CA 93955-6771

COST-EFFECTIVENESS ANALYSIS OF TRAINING IN THE DEPARTMENT OF DEFENSE

Henry Simpson

June 1995

DEFENSE MANPOWER DATA CENTER
Training and Readiness Evaluation and Analysis Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)

2. REPORT DATE
June 1995

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Cost Effectiveness Analysis of Training in the Department of Defense

5. FUNDING NUMBERS

6. AUTHOR(S)
Henry Simpson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Defense Manpower Data Center
DoD Center, Monterey Bay
400 Gigling Road
Seaside, CA 93955-6771

8. PERFORMING ORGANIZATION REPORT NUMBER
DMDC TR 95-004

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Deputy Under Secretary of Defense (Readiness)
4000 Defense, The Pentagon
Washington, DC 20301-4000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)
A study was conducted to determine the current state of knowledge and research on conducting the cost-effectiveness analysis of training (CEAT) in the Department of Defense based on a literature review, analyses, and a survey of subject matter experts. Findings were that (1) CEAT methods are inadequately defined, (2) DoD policy guidance for CEAT is ambiguous, (3) CEAT procedural guidance is inadequate, and (4) CEAT programs differ among the Services. CEAT is not a single method but a family of related methods. The cost analysis part of CEAT is fairly well defined but methods for performing the related training effectiveness analysis (TEA) are not. The study identified 16 different classes of TEA and no guidance for determining what type of TEA to perform under different conditions. Key DoD instructions are ambiguous about CEAT requirements and seem to exclude many training systems for which CEAT might be appropriate. There is no comprehensive guide on the conduct of CEAT, and existing procedural guidance is fragmented.

14. SUBJECT TERMS
Cost-effectiveness, training effectiveness analysis, cost analysis, military training, training systems, training development

15. NUMBER OF PAGES
77

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)

PREFACE

The Office of the Deputy Under Secretary of Defense for Readiness (DUSD[R]) has expressed concern that cost-effectiveness analyses of new training systems may often be performed poorly or not at all and that the Services may adopt systems without adequate justification. In response to its concern, the DUSD(R) requested that the Defense Manpower Data Center (DMDC) conduct a a study to determine whether DoD policy guidance or or perhaps other action is needed to facilitate more effective analyses in the Services. This report describes the work performed by DMDC in response to the DUSD's request.

ACKNOWLEDGMENTS

This report was reviewed by William West and Richard Evans of DMDC, Donald Johnson of DUSD(R), Jesse Orlansky of the Institute for Defense Analyses, John Boldovici of the U.S. Army Research Institute for the Behavioral and Social Sciences, and William Rankin of the Naval Air Warfare Center (Training Systems Division). Portions of the report were reviewed by Paul Sticha of the Human Resources Research Organization. The author is particularly indebted to Jesse Orlansky and John Boldovici for their thoughtul comments and suggestions for improving the report. The individuals listed below provided information, opinions, documents, and in other ways supported the study. The author thanks all for their help. Final responsibility for the report's content is the author's.

Dee Andrews	Air Force Armstong Laboratory
John Boldovici	U.S. Army Research Institute for the Behavioral and Social Sciences
Burke Burright	Air Force Armstrong Laboratory
Walter Butler	TRADOC Analysis Center
Richard Evans	Defense Manpower Data Center
Dorothy Finley	U.S. Army Research Institute for the Behavioral and Social Sciences
Eugene Hall	Independent Consultant
Donald Johnson	Office of the Deputy Under Secretary of Defense for Readiness
Peter Kincaid	University of Central Florida Institute for Simulation and Training
C. Mazie Knerr	Human Resources Research Organization
Vincent Lauter	Defense Manpower Data Center
Howard McFann[1]	Defense Manpower Data Center
Eugene Micheli	Naval Air Warfare Center, Training Systems Division
Robert Nullmeyer	Air Force Armstrong Laboratory
Jesse Orlansky	Institute for Defense Analyses
Michael Singer	U.S. Army Research Institute for Behavioral and Social Sciences
Paul Sticha	Human Resources Research Organization
Barbara Taylor	Navy Personnel Research and Development Center
Diana Tierney	Headquarters, TRADOC
Harold Wagner	U.S. Army Research Institute for the Behavioral and Social Sciences
Robert Witmer	U.S. Army Research Institute for the Behavioral and Social Sciences

[1] Deceased.

EXECUTIVE SUMMARY

Problem and Issues

The Department of Defense (DoD) invests heavily in training. The cost of individual training of military students for FY94 is approximately $14.2 B, contract expenditures for simulation and training for FY95 are estimated to be approximately $2.8B, and it has been estimated informally that collective training in operational units costs $40B to $50B each year. One way to leverage resources is to use training innovations (e.g., technologies, improved training methods) to increase training efficiency. Making the tradeoff among completing training alternatives is done using a class of methods that involve the cost effectiveness analysis of training (CEAT). The Office of the Deputy Under Secretary of Defense for Readiness (DUSD [R]) has expressed concern that CEAT may often be performed poorly or not at all and that the Services may adopt training systems without adequate justification. In response to its concern, the DUSD (R) requested that the Defense Manpower Data Center (DMDC) conduct a study to determine whether DoD policy guidance or perhaps other action is needed to facilitate more effective CEAT in the Services.

Objectives

Objectives of the study were to:

- Determine the current state of knowledge and research on conducting CEAT
- Identify documented CEAT methods
- Develop a CEAT general conceptual model
- Assess the current status of CEAT in the Services
- Determine potential areas where R&D on CEAT methods would be useful

Method

The method consisted of literature review, analyses, and survey of subject matter experts (SME).

Findings

CEAT Methods Are Not Well Defined

CEAT is not a single method but a family of related methods. The cost analysis part of CEAT is fairly well defined. However, performing the related training effectiveness analysis (TEA) poses at least two problems: (1) deciding what type of TEA to perform and (2) actually performing the TEA. Analyses suggested that there are 16 different classes of TEA. Hence, there are several times 16 ways to perform a TEA or CEAT.

Methods of collective training assessment are not fully developed. Conducting CEAT for systems intended to train groups of people remains difficult. More R&D needs to be performed to refine these methods.

Analytical CEAT methods hold out the promise of providing useful data in situations that preclude empirical methods, but the study revealed that (a) development of analytical methods has languished in recent years due to lack of resources, (b) methods are often perceived by users to be difficult to apply and to lack "user friendliness," (c) methods lack validation by comparison of their results with empirical methods, and (d) proponents often find it difficult to convince military decision makers that analytical methods produce valid results.

DoD Policy Guidance for CEAT Is Ambiguous

Key DoD instructions are ambiguous about CEAT requirements and seem to exclude many training systems that do not fit the definition of training device, simulator, or system (e.g., distance learning technologies, training delivery media) for which CEAT might be appropriate. The Army has published regulations making conduct of CEAT Army policy but the Navy and Air Force have not.

CEAT Procedural Guidance Is Inadequate

There is no comprehensive guide on the conduct of CEAT. Existing procedural guidance is fragmented. The complexity of CEAT precludes the development of a cookbook-style "how to" guide for conducting CEAT under all circumstances. It would be more realistic to assemble a set of CEAT resources that could be used in a modular fashion.

CEAT Programs Differ among the Services

The study defined a CEAT "program" in terms of a Service's published CEAT requirements, organization to perform CEAT, and publication of reports. Based on this definition, the Army has a CEAT program but the Navy and Air Force do not. CEAT is not performed in a consistent manner in the Navy and Air Force, although it may occur when the perception of need arises.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Problem and Issues

The Department of Defense (DoD) invests heavily in training. The *Military Manpower Training Report* (DoD, 1994) and the DoD budget (Clinton, 1994) indicate that the cost of individual training of military students for FY94 accounts for approximately 5.6% of the DoD budget ($14.2 B). The percentage of DoD personnel engaged in individual training as students and support staff ranges by Service from 14.1% to 20.4% of the total Service force, with an overall average in the DoD of 17.4% (Table 1). These figures do not include the cost of training on the job or within units, primarily because no DoD report provides this information. Total military contract expenditures for simulation and training for FY95 are estimated to be approximately $2.8B (Frost & Sullivan, 1994). It has been estimated informally that collective training in operational units costs $40B to $50B each year; the cost of on-the job training is unknown (Orlansky, 1994).

Table 1

Estimated Numbers and Percentages of Personnel Engaged in Individual Military Training as Students and Support Personnel by Service and Overall during FY94. Numbers represent thousands of personnel. (From *Military Manpower Training Report: FY 1995* (DoD, 1994).)

Service	Training cadre			Force size	Training as % of force size
	Load	Support	Total		
Army	54.2	56.0	110.2	540.0	20.4
Navy	45.1	33.0	78.1	471.5	16.6
USMC	18.0	14.0	32.0	174.0	18.4
USAF	29.9	30.0	59.9	425.0	14.1
Total	147.3	133.0	280.3	1610.5	17.4

The DoD has always needed to make efficient use of its training resources, and even more so in a time of downsizing and declining budgets. One way to leverage resources is to use training innovations (e.g., new technologies, improved training methods) to increase training efficiency. This makes sense if an innovation provides adequate training and costs less than the traditional training method. Making the tradeoff among competing training alternatives is done using a class of methods that assess the cost-effectiveness of training. For shorthand, this report refers to these methods collectively as cost-effectiveness analysis of training (CEAT).[1] The DUSD (R) has expressed concern that CEAT may often be performed poorly or not at all and that the Services may adopt training systems without adequate justification.

[1] Historically, the Army has used "CTEA" to refer to specific types of cost and training effectiveness analyses but the other Services have not generally used this terminology. CEAT is used here as a broad umbrella term encompassing all CEAT methods in all the Services.

The DUSD (R) requested that the DMDC conduct a study to determine whether DoD policy guidance or perhaps other action is needed to facilitate more effective CEAT in the Services. Some possible reasons why CEAT might not be performed well in the Services are:

- CEAT methods are inadequately defined
- DoD policy guidance is inadequate
- CEAT procedural guidance is inadequate
- Services lack adequate CEAT programs

This study was designed to gather information relating to these possibilities.

Objectives

Objectives of the study were to:

- Determine the current state of knowledge and research on conducting CEAT
- Identify documented CEAT methods
- Develop a CEAT general conceptual model
- Assess the current status of CEAT in the Services
- Determine potential areas where R&D on CEAT methods would be useful

Method

The method consisted of literature review, analyses, and survey of CEAT SMEs.

Literature Review

Most of the information presented in this report is based on the literature review. The analyses attempted to organize and in some cases integrate this information. CEAT SMEs (subject-matter experts) experts were consulted to validate the author's interpretations and conclusions. The literature review was conducted to identify the current state of knowledge and research on conducting CEAT. Documents were obtained, reviewed for relevance, and classified. Document content was analyzed to obtain answers to questions relating to project objectives. The literature review is discussed in greater detail below.

An electronic search was conducted of the Defense Technical Information Center database to identify documents produced since 1974 relating to training effectiveness, cost analysis, cost and training effectiveness, cost-effectiveness, and various combinations of these and related terms. SMEs suggested and in some cases provided additional documents. Another source of documents was the Training Effectiveness Catalogue System database generated during a recent project relating to collective training effectiveness assessment (Resource Consultants, Inc., 1992). The literature review covered several hundred documents, of which those listed in *References* were particularly useful to the project. In addition, a summary report of work unit, studies, and analysis efforts was obtained from the Manpower and Training Research Information System (MATRIS) on the subject of CEAT.

Much has been written about CEAT from many different perspectives. The largest part of this body of work probably consists of test reports, which the Army in particular has been prolific in publishing. Service organizations, research laboratories, and DoD contractors have published many studies in areas relating to CEAT (e.g., training effectiveness analysis, cost analysis). The DoD and the Services have published written CEAT guidance in the form of regulations, pamphlets, directives, and other documents that tell when analyses are required. There have been several attempts to provide "how to" guidance in the form of handbooks or similar documents aimed at the analyst. Several analytical methods have been developed and are described in technical reports, and periodic retrospective reviews have attempted to sort them out. Some analysts have written thoughtful papers over the years in attempts to refine CEAT methods as well as to point out their limitations. Meta-analyses have integrated CEAT and related work. Most of the foregoing literature was generated within the DoD community. Some non-DoD work is also relevant. CEAT has drawn from several threads of academic literature (e.g., cost-benefit analysis, measurement of training transfer, research design, meta-analytic methods). The literature review focused mainly on literature produced within the DoD community, but also included some academic literature.

Analyses

CEAT concepts and methods were reviewed, analyzed, and described in written form to develop a unified descriptive framework and CEAT conceptual model. These analyses led to the information presented in the report sections titled *CEAT Concepts* and *CEAT Methods.*

DoD and Service-written documentation relating to CEAT were analyzed to determine scope and adequacy. This analysis led to the information presented in the report section titled *CEAT Written Guidance.*

Survey of Subject-Matter Experts

CEAT SMEs were surveyed to gather information on the status of CEAT in the Services and to answer questions that arose during the study. Most of the SMEs had made significant contributions to the CEAT literature and had first-hand knowledge about the conduct of CEAT in the Services. (Participating SMEs are listed in *Acknowledgments.*) Several SMEs were also surveyed to gather information on the status and use of analytical CEAT methods. The survey included telephone interviews with SMEs representing or familiar with CEAT in the Army, Navy, and Air Force. Discussion points to be covered during the interview were listed in a protocol. The protocol was faxed to SMEs prior to the interview to enable them to prepare. It was used during the interview to insure that essential discussion points were covered, but was not adhered to rigidly; many of the interviews expanded to cover topics in the SME's particular area of expertise or interest. In addition, several SMEs were contacted informally during the study to discuss issues and answer questions. The report section titled *CEAT in the Services* is based in part on this survey.

Report Overview

The report is organized in six sections, each focused on one or more of the study's objectives. *Introduction* describes the problem and issues, study objectives, method, and provides a report overview. *CEAT Concepts* sketches several important concepts relating to the conduct of CEAT in the DoD. It discusses cost-benefit analysis, cost-effectiveness analysis, cost and operational effectiveness analysis, and CEAT. It also presents a CEAT general conceptual model. *CEAT Methods* discusses the impact of time on CEAT, presents a taxonomy of TEA methods, describes several different empirical and analytical TEA methods as they relate to the taxonomy, and discusses cost analysis and sensitivity analysis. *CEAT Written Guidance* reviews CEAT written guidance provided by the DoD, Army, and other sources. *CEAT in the Services*, provides an overview of how the Army, Navy, and Air Force deal with CEAT. *Conclusions* summarizes the study's analyses and findings and identifies potential OSD actions that might be useful.

CEAT CONCEPTS

This section broadly sketches a set of concepts relating to the conduct of CEAT in the DoD. CEAT terminology can be confusing, particularly if one has not studied it closely. Hence, this section provides background information and context for the rest of the report. The section begins with a discussion of cost-benefit analysis (CBA), which is used to make cost-benefit decisions in the public sector. Cost-effectiveness analysis (CEA), the military's equivalent of CBA, is then described. The next two subsections describe two specialized forms of CEA: cost and operational effectiveness analysis (COEA) and CEAT. The first three subsections provide fairly brief summaries of the analytical techniques. The CEAT subsection is developed at a greater level of detail as it deals with the main subject of the report.

Cost-Benefit Analysis

Definition

CBA is used in the public sector to make decisions regarding alternative courses of action where the inputs and outcomes (benefits) can be expressed in dollar terms; these have implications for societal welfare and the allocation of public funds (McMichael, 1985). Examples of such decisions are choosing among a set of alternative (a) water treatment plant designs, (b) health care systems, (c) procedures for recruiting and retaining police officers. Sassone and Schaffer (1978) define CBA as "an estimation and evaluation of net benefits associated with alternatives for achieving defined public goals" (p. 2). The definition tells several things about CBA. First, CBA is used to help meet public goals. Second, CBA compares alternative courses of action rather than evaluating a single, chosen course. Third, a process ("estimation and evaluation") is used to make the comparison. Fourth, certain criteria ("net benefits") are used to decide the outcome.

Costs and Benefits

McMichael contrasts CBA, which is used in the public sector, with profitability analysis, which is used in business and industry. While profitability analysis attempts to maximize profits, CBA takes the broader societal view of both costs and benefits. The objective of CBA is to increase benefits to society in terms of economic efficiency. CBA requires that it be possible to express benefits in terms of cost (Derrick & Davis, 1993).

Costs of alternatives are typically estimated directly using cost models that take into account all of the associated costs of the alternatives throughout a projected life cycle.

To make comparisons among alternatives, benefits must be expressed in terms of cost. McMichael states that benefits in CBA are normally valued based on willingness to pay by the public. Value can also be estimated using several other techniques.

Process

CBA encompasses a wide range of procedures and is not a single technique. Sassone and Schaffer contend that though CBA incorporates certain general principles, it is difficult if not impossible to design an all-purpose CBA procedure because of differences in public projects. They provide a basic framework for conducting a CBA consisting of initial planning stages followed by data collection, separate cost and benefit analyses, and presentation of results. McMichael provides a similar framework. While there are some differences in their formulations, the authors would probably agree with Swope (1976) that a CBA process should include the following steps:

- Formulate Assumptions
- Determine Alternatives
- Determine Costs and Benefits
- Compare and Select Alternatives
- Conduct Sensitivity Analysis

Assumptions are usually made regarding what variables will affect the process and the range of values those variables will present. The alternatives will include the new system and one or more other possibilities. Frequently, one of these is an existing system. After the costs and benefits of alternatives have been determined, they are compared and a selection is made. In CBA, the best alternative is the one yielding the greatest net benefit (i.e., the alternative whose benefit value (expressed in monetary terms) less its cost is the greatest). Orlansky (1989) provides the following concrete example:

> [In] cost-benefit analysis...both the input and output values can be measured in monetary terms. This requires an open market to assess the value...of the output that results from a particular use of resources (i.e., the costs). One example might be a cost-benefit analysis of a particular form of advertising. The costs are those

needed to develop and conduct a particular advertising program; the benefits are the profits that may be attributed to the advertising program (p. ix).

Assumptions are required in planning a CBA and these can lead to uncertainty in the outcomes of analyses. If the CBA is locked into a single set of assumptions with the intent of obtaining a definitive result, its outcome may be too fragile to be trustworthy. It is more sensible to vary the assumptions systematically and to provide the results of analyses under different assumptions. This procedure is referred to as sensitivity analysis.

In outline, this process seems simple, though in application it is complex and readers should consult the cited works for details. To conclude discussion of CBA, let us reiterate with slight elaboration the points made above: (1) CBA is used to meet public goals, (2) it compares alternatives, (3) a process is used to make the comparison, (4) cost and benefit criteria (both in monetary terms) decide the outcome.

Cost-Effectiveness Analysis

Definition

Cost-effectiveness analysis (CEA) is the method used in the DoD to make decisions regarding alternative courses of action where the outcomes affect military performance. In these cases, there is no market available to establish the monetary value of the output (performance) although inputs can be expressed in monetary terms. Examples are choosing among a set of alternative (a) weapon systems, (b) weapon system upgrade programs, (c) training methods. A definition of CEA analogous to that given earlier for CBA might be *an estimation and evaluation of the military value associated with alternatives for achieving defined military goals.* CEA is used to help meet military goals (rather than CBA's public goals). CEA, like CBA, compares alternatives using a formal process. Criteria decide the outcome for both CEA and CBA, but the criteria differ (i.e., military value for CEA and public benefits for CBA). *Economic analysis*, a term used in a number of DoD publications, has a meaning synonymous with CEA (Rankin & Swope, 1991).

Costs and Military Value

Costs of alternatives in a CEA are estimated in a manner similar to that of a CBA by using cost models that take into account all of the associated costs of the alternatives throughout a projected life cycle.

However, estimating military value for a CEA is different from estimating public benefits in a CBA. An important difference between CEA and CBA is that the outcome (military value) is not defined in the same terms as cost (Orlansky, 1989; Rankin & Swope, 1991). Orlansky commented on this matter as follows:

> [The cost-benefit] procedure cannot be followed when examining the products of a military weapon or training program. There is no open market that can establish

the monetary value of increased readiness, better trained personnel, or better weapons (p. ix).

Ultimately, military value is reflected in the degree of combat success. Weapon system A has greater military value than weapon system B if A is more likely to prevail in battle than B. Or, if two training alternatives are being compared, treatment A has greater military value than treatment B if A better equips students to prevail in battle than B. Military value can be assessed empirically only in combat and it is impractical to wait for a war to make the assessment. An alternative to combat is to create a combat-like environment (e.g., to use an instrumented live exercise). In performing CEA, measures of effectiveness (MOE) are used which ostensibly predict combat success. (In the experimental paradigm, MOE are equivalent to dependent variables, the variables used to assess the impact of an experimental treatment condition.) Some of the assumptions and potential problems of using surrogate measures are discussed in greater detail in the CEAT subsection. The concept of military value of training is developed in detail in Gorman (1990) and Deitchman (1990).

Process

Like CBA, CEA encompasses a wide range of procedures and is not a single technique. Because of conceptual similarities between CBA and CEA, it is reasonable to extend Sassone and Schaffer's contention regarding the difficulty of designing an all-purpose CBA procedure to the realm of CEA. Likewise, the basic framework for conducting a CEA parallels that of a CBA, described by Swope (1976), but with a slight change to the third step ("Benefits" becomes "Military Value"):

- Formulate Assumptions
- Determine Alternatives
- Determine Costs and Military Value
- Compare and Select Alternatives
- Conduct Sensitivity Analysis

Since cost and military value use different units, selection of alternatives cannot be done on a cost basis alone as with CBA. Orlansky has described the decision-making logic in several published papers (e.g., Orlansky, 1985, 1989, 1990) (Figure 1). Though the logic was developed in the context of CEAT, it generalizes to CEA.

COST	**EFFECTIVENESS**		
	LESS	**SAME**	**MORE**
LESS	UNCERTAIN	ADOPT	ADOPT
SAME	REJECT	UNCERTAIN	ADOPT
MORE	REJECT	REJECT	UNCERTAIN

Figure 1. Orlansky's decision logic diagram for evaluating the relative effectiveness and cost of two training methods during CEA.

Orlansky (1989) commented as follows on the interpretation of the diagram:

> a. If one alternative is as effective or more effective than another and it costs less, adopt it; it is also the preferred choice if it is more effective and costs the same.
> b. If an alternative is less effective and costs the same or more than another to which it has been compared, reject it; this is also the case if it is equally effective but costs more.
> c. If any of the following combinations of the cost and effectiveness of an alternative is found, no rational preference can be made: (1) less effective and less cost; (2) equal effectiveness and equal cost; (3) more effective and more cost. (pp. xiii-xiv.)

As with CBA, discussions of CEA at this level seem simple though actually performing an analysis is more complex. To conclude discussion of CEA, let us reiterate with slight elaboration the points made above: (1) CEA is used to meet military goals, (2) it compares alternatives, (3) a process is used to make the comparison, (4) cost and military value criteria decide the outcome.

Cost and Operational Effectiveness Analysis

Cost and operational effectiveness analysis (COEA) is the specific form of CEA used in the DoD to make decisions regarding alternative courses of action for materiel systems. *DoD Directive 5000.2: Defense Acquisition Management Policies and Procedures* establishes policies and procedures for the conduct of COEA primarily for the purpose of supporting milestone decision reviews. *DoD Instruction 5000.2M, Defense Acquisition Management Documentation and Reports* (DoD, 1991), states that a COEA "evaluates the costs and benefits (i.e., the operational effectiveness or military utility) of alternative courses of action to meet recognized defense needs (p. 8-1). At a conceptual level, COEA is a type of CEA so the definition of CEA given earlier applies to COEA as well and is more general.

Costs of alternatives in a COEA are estimated as for a CEA, taking into account all costs associated with the alternatives throughout a projected life cycle. According to *DoD Instruction 5000.2,* life cycle cost reflects the cumulative costs of developing, procuring, operating, and supporting the system.

Operational effectiveness is assessed using MOE. This is what *DoD Instruction 5000.2* says about MOE:

> [MOE] should be defined to measure operational capabilities in terms of engagement or battle outcomes. Measures of performance, such as weight and speed, should relate to the [MOE] such that the effect of a change in the measure of performance can be related to a change in the [MOE] (p. 4-E-3).

DoD Instruction 5000.2M adds that MOE show how well alternatives meet functional objectives and mission needs and offers as examples loss exchange ratios, force effectiveness contributions, systems saved, and tons delivered per day. The intent is to determine military value as reflected in engagement or battle outcomes, though in practice this can be exceedingly difficult.

Conceptually, the COEA process *is* the CEA process, discussed earlier. The third step is changed to express military value in terms of operational effectiveness:

- Formulate Assumptions
- Determine Alternatives
- Determine Costs and Operational Effectiveness
- Compare and Select Alternatives
- Conduct Sensitivity Analysis

Cost-Effectiveness Analysis of Training

Overview

CEAT is the specific form of CEA used in the DoD to make decisions regarding alternative courses of action for training. Examples are choosing among a set of alternative training (a) methods, (b) simulators, (c) devices. A definition of CEAT analogous to that given earlier for CBA might be *an estimation and evaluation of the training effectiveness and costs associated with a set of training alternatives.* Sassone (1985) defines CEAT as "comparison of the effectiveness and costs of alternative training systems" (p. 2). TRADOC defines CEAT as a "process that assesses the variable effectiveness and variable costs associated with a set of alternative training subsystems" (Department of the Army, 1980, p. 2-2). These all say much the same thing. And, at a conceptual level, CEAT is a type of CEA, so the definition of CEA given earlier also applies to CEAT.

Conceptually, the CEAT process *is* the CEA process, discussed earlier. The third step is changed to express military value in terms of training effectiveness:

- Formulate Assumptions
- Determine Alternatives
- Determine Costs and Training Effectiveness
- Compare and Select Alternatives
- Conduct Sensitivity Analysis

Costs of alternatives in a CEAT are estimated as for a CEA, taking into account all costs associated with the alternatives throughout a projected life cycle.

Training effectiveness is assessed using training effectiveness analysis (TEA).

The foregoing is the CEAT process in overview.

Before moving on, it may be useful to reprise the similarities and differences of the four types of analyses sketched in this subsection. Table 2 compares CBA, CEA, COEA, and CEAT in terms of alternatives being compared, criteria used, and the type of decision-making process used.

Table 2

Comparison of CBA, CEA, COEA, and CEAT in Terms of Alternatives Being Compared, Criteria Used, and Decision-Making Process Used

	CBA	CEA	COEA	CEAT
Alternatives	Public policies, procedures, etc.	Any military system	Weapon systems	Training systems
Criterion 1	Cost	Cost	Cost	Cost
Criterion 2	Benefits	Military value	Operational effectiveness	Training effectiveness
Decision making	Best net benefits (benefits - cost)	Best combination of Criterion 1 & 2	Best combination of Criterion 1 & 2	Best combination of Criterion 1 & 2

The remainder of this section describes the elements of CEAT in greater depth with subsections covering training effectiveness analysis and cost analysis.

Training Effectiveness Analysis

There is an enormous literature on TEA and attempting to comprehend any part of it is challenging. TEA is difficult because training can occur in many different contexts, be of different types, and deciding what and how to measure is seldom obvious, to name a few of the problems. Some key issues in TEA are discussed below.

Training Environment. Training environment is *where* training occurs. In the military, the distinction is usually made between training received in schools (sometimes called institutional training) and training received in units. In general, institutional training is structured and often occurs in a classroom or laboratory setting. Unit training occurs in the unit setting, often using actual equipment. Simulators and training devices are used in both settings.

Individual and Collective Training. The distinction in the heading is between individuals and groups of people (i.e., *who* receives training). Individual training is training given to individual members of the military to develop their skills. Such training is based on individual tasks. An example of an individual task would be to troubleshoot an electronic circuit. Formally structured military training is usually defined in terms of tasks (DoD, 1990), with their associated conditions and performance standards. Individual training is provided both at schools and on the job. The cost of the former is reasonably well known; the cost of the latter is not known except for some estimates in a few studies.

Collective training is given to groups of individuals (e.g., crews, teams, units) who must work together and coordinate their activities. Collective training is defined in terms of collective tasks (examples given below). While it is convenient to divide training into two categories, this is an oversimplification, for there is more accurately a hierarchy of collectiveness. Deitchman (1993) outlines a broad four-level hierarchy of military missions, with each mission level encompassing those below. In his example, the highest level is to *win a war*; below, successively, are *succeed in battle*, *operate a military unit*, and *engage the enemy*. Sassone (1985) develops a multi-level battalion effectiveness hierarchy in which battalion effectiveness is reflected in resources, training programs, and proficiencies at successively higher levels in the battalion hierarchy (e.g., individual, squad, platoon, company, battalion.)

Team training is a type of collective training involving relatively small groups whose hierarchy, if it exists at all, is limited. Salas, Dickinson, Converse, and Tannenbaum (1992) define a team as:

> a distinguishable set of two or more people who interact, dynamically, independently, and adapatively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span of membership (p. 4).

To a degree, this definition applies more generally to collective training. However, the focus of the definition is on a small group of people who work closely together throughout the life of the team. A collective may include many different teams whose interactions with one another are not as intimate or as continuous across time. Turnage, Houser, and Hofmann (1990) make the following important distinctions between collective and team training research:

> ..."collective" performance assessment research has not been conducted extensively to date; most research has focused on "teams" as the unit of measurement. Although much small group (team) training research may generalize to larger units (e.g., corps, division, brigade, battalion, platoon), the implicit assumption is that small group (team) research is more productive because it is "cleaner" from both a conceptual and measurement standpoint.... [The] terms "team" and "collective" are not synonymous (p. 1-1).

These descriptions of how one may break down individual and collective training illustrate that the process is less than straightforward. This suggests the difficulty of defining suitable MOE to use in TEA, particularly for collective training.

Training Taxonomy. Gorman (1990) observed that training environment and type of training (individual vs. collective) cross to form a four-element taxonomy:

> Training conducted by an armed force to prepare its members for war occurs in four regimes differentiated by the target (object) of the training--whether individuals or collectives--or its environment (venue)--whether in institutions or in units; in short, defined by who is being trained and where the training is taking place (p. 23).

Figure 2 illustrates Gorman's taxonomy. Concrete examples of each of the four classes of training are given in the cells of Figure 2. The taxonomy helps illustrate the many ways that training can occur. And, since one may have to direct a CEAT at any of these ways of training, the taxonomy is relevant.

		Training Environment	
		School	**Unit**
Type of Training	**Individual**	1. Participation in a course at a resident service school, learning to troubleshoot with a maintenance simulator, flight training	2. Participation in a class conducted by a supervisor, supervised on-the-job training, practicing tank gunnery using operational equipment
	Collective	3. Naval damage control training, tank crew drills, formation flying	4. Unit training in the field, unit training with networked simulators, joint and coalition exercise.

Figure 2. Training taxonomy described by Gorman (1990) illustrating four training regimes of military training.

Angier, Alluisi, and Horowitz (1992) expanded Gorman's taxonomy to incorporate various types of simulators (Figure 3), and no doubt there are other ways of sorting out how training may occur in different contexts. For example, if one differentiated between unit training and combat, a third column might be added to Figure 2 for the individual and collective training that occurred as the result of combat experience.

Measures of Effectiveness.[2] CEAT MOE are used to make comparisons among training alternatives. *DoD Instruction 5000.2M* defines MOE as "tools that assist in discriminating among a number of alternatives. They show how the alternatives compare in meeting functional objectives and mission needs" (p. 8-7). For example, one way to make comparisons is to conduct an experiment in which two training treatments are given, MOE data are collected, and the treatment with the best MOE scores wins the competition. In this example, using the experimental paradigm, MOE perform the role of dependent variables.

---

[2] The literature relating to CEAT MOEs is fragmented. There is no single document that deals with the subject comprehensively. Barr (1986) provides good coverage of MOEs for use in systems analysis that might serve as a model in CEAT. Among other things, Barr provides a multi-level hierarchy of MOEs, MOE development procedures, and sources of published MOEs and related performance measures. Elsewhere, guidance in the literature on how to select or create MOEs is general and vague. For example, *TRADOC Pamphlet 11-8: Army Programs Studies and Analysis Handbook,* advises readers that "selection of the MOE is a subjective process based on how the study agency believes force effectiveness may be best assessed" (p. 2-3).

<table>
<tr><td colspan="2"></td><td colspan="3">Training Environment</td></tr>
<tr><td colspan="2"></td><td>School</td><td>Continuation</td><td>Unit</td></tr>
<tr><td rowspan="2">Training Focus</td><td>Individual</td><td>Maintenance</td><td>Flight</td><td>Embedded Training, portable part-task trainers</td></tr>
<tr><td>Collective</td><td colspan="2">Wargaming, crew training (e.g., C-130)</td><td>Embedded training, networked simulators</td></tr>
</table>

Figure 3. **Expansion of Gorman's training taxonomy to incorporate various types of simulators. (From Angier et al., 1992.)**

The Transfer Assumption. Earlier discussion of CEA made the point that MOE ostensibly predict combat success and alluded to the fact that this indicated certain assumptions and gave rise to potential problems. For example, if one is conducting a TEA to compare two different forms of individual training in school, the obvious (if not necessarily best) choice of MOE would be student grades in school. After conducting courses using both forms of training, it would be reasonable to identify the best form of training based on student grades. The assumption being made here is that school grades bear some relationship to combat success.[3]

Actually, the line of reasoning implies a chain of assumptions; that is, that school performance affects job performance, which in turn affects combat readiness, which in turn affects combat performance (Solomon, 1986). At each link of this chain, an indicator is used as the surrogate for the next link. Orlansky (1989) reports that a small number of studies (e.g., Orlansky, 1985; Hammon & Horowitz, 1987; Gibson & Orlansky, 1986) provide robust data supporting the linkage between performance in school, on the job, in field exercises, and later on to military readiness, but that these linkages have been largely uninvestigated.

Using surrogate measures has risks but, as a practical matter, in conducting a CEAT it seems unavoidable. It is important to select the MOE that best predict combat success and to acknowledge the limitations of whatever MOE are chosen.

Quality Distinctions among MOE. Several distinctions are commonly made in TEA that affect the selection of MOE. One distinction is between internal and external evaluation (e.g., Department of the Army, 1988b). Internal evaluation focuses on training processes. Hall,

[3] The formal name for this is transfer of training (or transfer of learning), which may be thought of as the degree to which training in situation A prepares one to perform in situation B. Transfer is discussed in greater detail later.

Rankin, and Aagard (1976) give as examples clarity of training content, quality of training aids, and media available. ("Internal evaluation" and "process evaluation" are synonymous.) External evaluation focuses on training products (e.g., student performance at the conclusion of training.) ("External evaluation" and "product evaluation" are synonymous.) A TEA may include both types of evaluations. Hall et al. (1976) caution that process data may provide information useful for improving training but do not assess training effectiveness. To do that, product data must be collected.

Another distinction is between subjective and objective data. Subjective data are based on the opinions, judgment, and wisdom of people who generate data during a TEA. An example would be an observer's rating of how well an instructor delivered a demonstration during a class. Objective data are based on observable events whose occurrence or non-occurrence is not usually subject to dispute. An example would be a student's accuracy score on an end-of-course performance test. Objective data generally have greater face validity. However, subjective data increase in value as the situation becomes less structured. Deitchman (1993) contends that it is particularly important to use expert judgment to capture intangibles such as leadership, motivation, morale, and the personality of the commander. An example of where this would apply is in evaluating a ship crew's performance during a combat simulation. In such cases, the evaluations of senior commanders and other SMEs carry great weight, although they may or may not be accurate and valid.

Some other important MOE quality factors are reliability, validity , unobtrusiveness, sensitivity, and practicality (Hall et al., 1976; McFann, 1983; Waag, Pierce, & Fessler, 1987).

MOE for Individual Training. There is no definitive guidance for developing MOE for individual training. What is presented here is a version, derived from several sources, that conveys the essence of what is involved in developing MOE.

The first requirement for a CEAT is a task list (Matlick, Berger, Knerr, & Chiorini, 1980). The availability of task data varies with the stage in the weapon system acquisition process (WSAP), with little data at the start and more as the system matures (DoD, 1991b). MOE for individual training assess performance on the tasks an individual is required to perform. If training is structured, whether in school or unit, MOE might start with the tasks, conditions, and standards in the training syllabus or plan. It may be that no task list exists or that the list is incomplete or inadequate. A deficient task list is most likely for a new training system, but may exist with an established one. If the tasks have not been fully defined, it may be necessary to perform a task analysis. Boldovici and Kraemer (1975) emphasize the importance of SME participation in this process to insure that it yields a comprehensive task list. Whatever the case, an adequate task list must be created. Tasks are then selected from the complete list which reflect the particular interests of the TEA. MOE are developed from the selected tasks.

MOE for Collective Training. The development of MOE for collective training appears to be more an art than a science. In a recent review of the state of the art in collective training evaluation, Turnage, Houser, and Hofmann (1990) commented as follows:

> The Army has long recognized that the performance of integrated crews, teams, and units is essential to overall mission success. Despite this, the current state of collective training evaluation has remained at a relatively unsophisticated level. Lack of understanding of the important dimensions of collective training and evaluation has hampered attempts to adequately assess combat readiness (p. iii).

The research in this area, particularly for team training, has made much progress in recent years, but can offer only general guidance on what tasks to target when evaluating collective training effectiveness. Another study, based on an extensive literature review and the inputs of an expert panel, concluded that there are "no ... universally accepted [MOE] that...relate to...collective training programs" (Resource Consultants, Inc., 1993, p. 54). As in the discussion of individual training MOE, what is presented is a hypothetical process based on several sources that is intended to convey the essence of what is involved in developing MOE.

As with MOE for individual training, the first requirement is a task list (Turnage et al.). Tasks are then selected from the list which reflect the particular interests of the TEA. MOE are developed from the selected tasks. This seemingly simple process is complicated by the fact that collective tasks tend to exist in hierarchies. A separate set of tasks can be defined for each level in the hierarchy. As the number of levels increases, additional tasks are added for each level and for the possible interactions of that level with other levels in the hierarchy (Figure 4). To complicate matters further, the nature of the hierarchy can vary depending upon preferences. For example, of the two collective task hierarchies described earlier, Deitchman's contains a few deep levels that do not correspond exactly to organizational structure, whereas Sassone's incorporates each level explicitly in its structure. McFann (1990) uses a functional categorization scheme whose categories do not correspond literally to organizational elements. For example, the measurement system he describes collects data on (among other things) critical combat functions such as C2, intelligence, maneuver, fire support, and air defense.

Hypothetically, a comprehensive task list can be created by defining all tasks at each level and all tasks involving interactions among levels. The comprehensive task list can then be pared down based on the particular interests of the TEA. Tasks remaining at the end of this process can be used as MOE. The process in summary:

- Define hierarchy (or other structure)
- Define tasks at each level
- Define tasks for interactions among levels
- Select tasks
- Create MOEs

The hierarchy or other structure may contain a single level (e.g., a team) or multiple levels (e.g., an armored battalion with its companies, platoons, squads). The process/product distinction made earlier is particularly important when dealing with collective MOE. The product MOE for a military unit is usually the most important indicator of unit performance; that is, accomplishing the mission (O'Neil, Baker, & Kazlauskas, 1990). (Many other factors also contribute to mission accomplishment, but in CEAT these factors are held constant and the MOE reflect the contribution of collective training.) McFann (1990) provides an excellent example of an existing measurement system that incorporates process and product measures. Figure 5 illustrates the ARI's (U.S. Army Research Institute for Behavioral and Social Sciences) Analytic Unit Performance Measurement System, which is used at Combat Training Centers. Process measures (left) are performance of tasks for seven operating systems and three battle phases. Product ("outcome") measures are mission results as reflected in Army standard METT-T (mission, enemy forces, troops friendly, terrain control, time) factors.

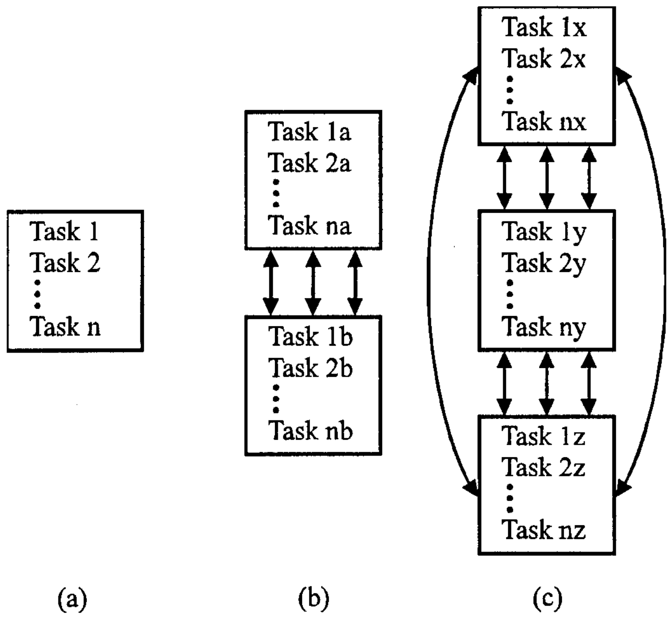


Figure 4. **Illustrating the proliferation of tasks as the number of levels in an organizational hierarchy increases: (a) with no hierarchy, analysis need only consider tasks for a single team; (b) with 2-level hierarchy, analysis must consider tasks at each level and additional tasks for interactions between levels; (c) with multi-level hierarchy, analysis must consider tasks at each level and tasks for all possible interactions among levels.**

McFann describes the analysis that led to this decomposition of unit tasks in terms of input, output, and process with the intent of determining system effectiveness and efficiency. The approach seems to derive not from a theory of collective training so much as from a production model.

Collective training is analyzed from a narrower perspective by those focusing on teams. The unit of analysis of this work is the small group. The processes which occur in teams reflect the application of team skills (e.g., communication, coordination, integration, self-evaluation, team awareness, and decision making) (Turnage et al.). Salas, Dickinson, Converse, and Tannenbaum (1992) provide an overview of several theoretical models that have been used in team research (e.g., normative, time and transition, task group effectiveness, team evolution and maturation, team performance, task orientation). They integrated the separate models into a general model (Figure 6), about which they stated the following:

> ...team performance is the outcome of dynamic processes reflected in the coordination and communication patterns that teams develop over time. The processes are influenced by organizational and situational characteristics, and task and work characteristics, as well as individual and team characteristics (pp. 15-16).

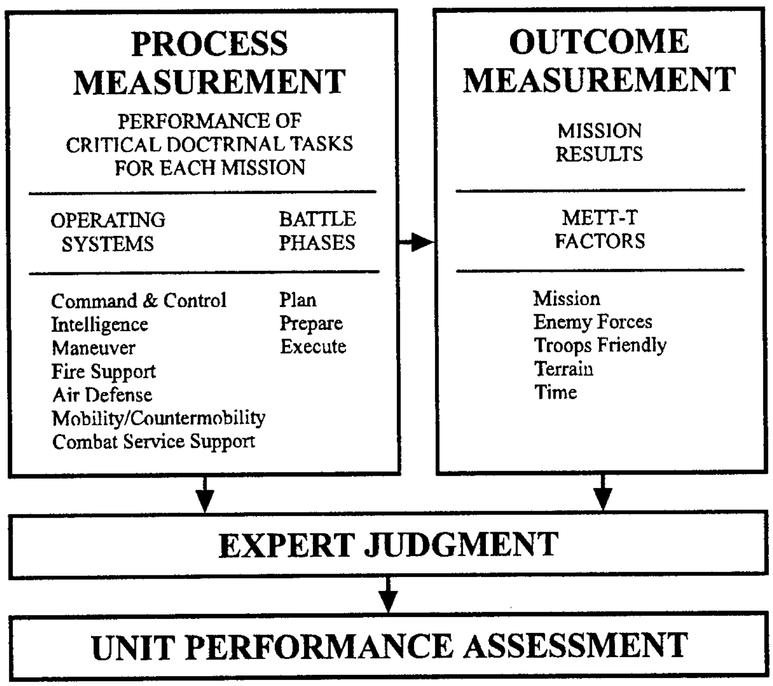


Figure 5. **ARI's Analytic Unit Performance Measurement System, which is used at Combat Training Centers. Process measures (left) are performance of tasks for (a) seven operating systems and (b) three battle phases. Product ("outcome") measures are mission results as reflected in Army standard METT-T factors (mission, enemy forces, troops friendly, terrain control, time). (Adapted from McFann, 1990.)**

Although the framework includes input, process ("throughput"), and output, it appears that the primary focus in on the internal workings of the team (i.e., team process). Elsewhere, Swezey and Salas (1992) declare that "the domain of teamwork deals with *process* issues" (p. 222) and provide a classification system for team processes meant to help in the design of team training (Table 3). Though many of the 12 categories apply exclusively to training, others apply to collective TEA (e.g., leadership, communication, adaptability, coordination and cooperation.) Hence, this framework could be used to build MOE to assess targeted team processes.

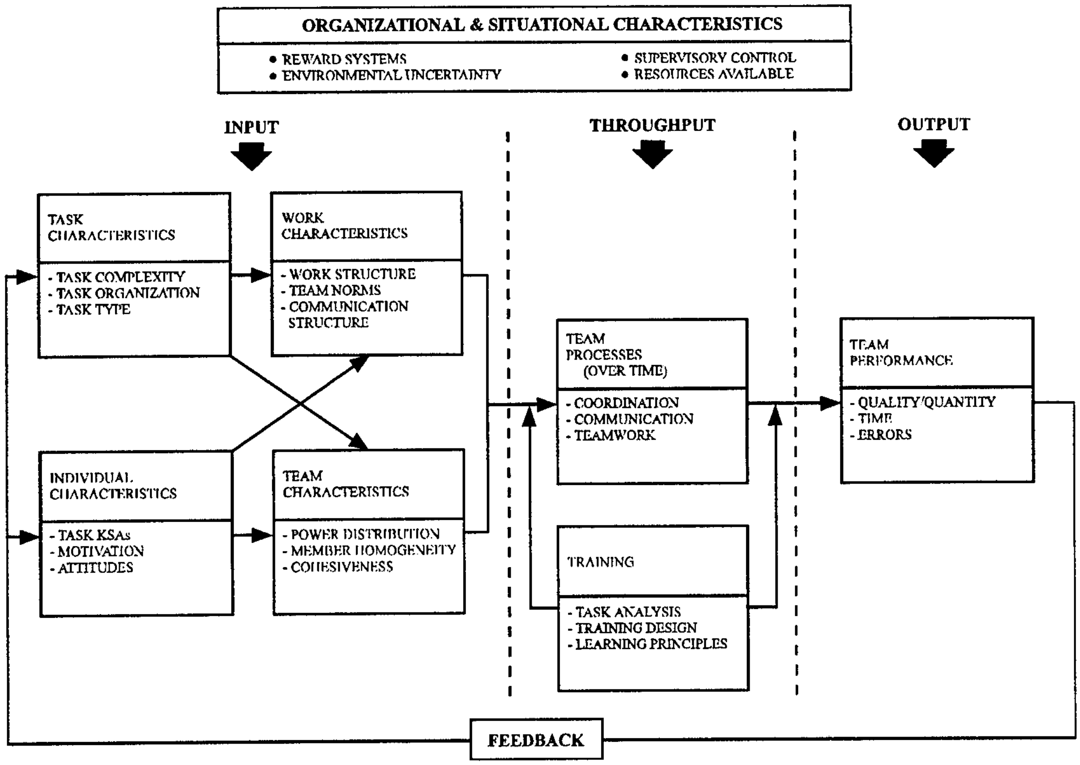


Figure 6. **Integrated model of team performance and training. (From Salas et al., 1992.)**

There is no formula for developing collective MOE. Hence, it is no surprise that many authors stress the importance of getting SMEs to participate in the development of MOEs. For example, Tuttle and Weaver (1986) describe a structured procedure to form teams to work together to identify and screen indicators to assess the productivity of Air Force organizations--a procedure that probably could be applied more generally (e.g., to defining product MOEs for combat organizations.) Boldovici and Kraemer (1975) stress the importance of using SMEs to review task analyses to insure that nothing is left out. And McFann (1990) emphasizes the importance of using the expert judgment of observer/controllers to obtain and interpret unit performance data.

It is always reasonable to call on the experts for their advice, for who knows better how to assess a complex situation? On the other hand, it can reasonably be argued that such reliance indicates that the essence of what the SMEs are judging has not been adequately captured in objective form. Hence, it remains an art rather than a science. O'Neil et al. contend that research is needed to lead to team theory which would enable better decomposition of team performance and measurement. To the extent that theory is developed and validated, the mystery of what teams do is reduced, and the matter moves into the realm of science rather than art.

Table 3

Classification System for Team Processes. (From Swezey & Salas, 1992.)

Element	Team Process Category
I	Team Mission and Goals
II	Environment and Operating Situation
III	Organization, Size, and Interation
IV	Motivation, Attitudes, and Cohesion
V	Leadership
VI	Communication: General, Conveying Information, Feedback
VII	Adaptability
VIII	Knowledge and Skill Development
IX	Coordination and Cooperation
X	Evaluation
XI	Team-Training Situation: General, Role of the Instructor, Training Methods
XII	Assessment of Team-Training Programs: Pretraining Assessment, Overall Assessment

Despite the fragmented nature of the area, some generalizations can be made about MOE for collective training:

- Tasks tend to exist in structures (e.g., hierarchies, functional categories)
- The structure requires definition
- Tasks at each level require definition
- Interactions among levels require definition
- Tasks should be selected which focus on the interests of the TEA
- Tasks are converted to MOE
- Product MOE (mission accomplishment) are the most important
- Process MOE assess the internal workings of the collective

Transfer. In studying transfer of training, "one is interested in the effect of a specifiable prior activity upon the learning of a given test activity" (Osgood, 1949, p. 432). Transfer reflects the effect of old learning in a new situation. Transfer can be positive, negative, or indeterminate. Positive transfer is obviously desirable and negative transfer undesirable, while indeterminate transfer indicates that training value is unknown. The conditions under which students learn (e.g., in the classroom with an instructor, with a training device or simulator) typically differ from those in which students apply the knowledge and skills on the job. Hence, transfer of training should concern anyone who cares about training effectiveness.

Study of transfer has produced a "voluminous literature" and a variety of ways to express the phenomenon in quantitative terms (Gagne, Foster, & Crowley, 1948). Osgood's explanation of the mechanism of transfer is based on a form of learning theory which expresses learning and performance in terms of stimulus-response pairs required in the "old" and "new" situations. The amount of transfer depends on similarity of stimulus and response between the two situations. Maximum transfer occurs with identical stimulus and response. Minimum transfer occurs with identical stimulus and different response because of interference. The amount of transfer varies between these two conditions depending upon relative amounts of similarity. (There are many other--and more current--theoretical explanations for transfer, though Osgood's continues to be cited and provides a useful framework for explaining the phenomenon.)

To illustrate how this theory might apply in an actual situation, assume that one wanted to predict transfer from a simulator to a piece of operational equipment such as a radar display. Both simulator and radar present stimuli in the form of displayed information and both require the operator to make responses (e.g., detect, track, and report targets). Osgood's theory predicts maximum transfer when simulator and radar present identical information and require identical operator responses. Theory predicts minimum transfer with identical information but different responses between simulator and radar.

It is important as a practical matter during TEA to measure the amount of transfer. Transfer formulas have traditionally been used for this purpose. The simplest transfer formula compares performance between an experimental and control group on a transfer task. The formula was presented by Roscoe (1971, 1972) in the following form:

$$\frac{Y_C - Y_X}{Y_C}$$

Y_C = time, trials, or errors required by a control group to reach a performance criterion
Y_X = corresponding value for an experimental, or transfer, group having received prior practice on another task

To illustrate, suppose one ran an experiment to determine how well a radar simulator worked during training as a substitute for an actual radar. Hypothetical data are shown in Table 4.

Table 4

Data from Hypothetical Training Transfer Experiment

Group	Hours on simulator	Hours on radar to reach criterion
Experimental	300	200
Control	0	400

In this case Yc = 400 and Yx = 200; substituting values yields:

$$\frac{400 - 200}{400} = 50\%$$

Roscoe and others noticed that the ratio is insensitive to the cost of time invested in the experimental condition (e.g., time on simulator).[4] So long as the time to reach criterion in the experimental condition remains the same, the formula will show the same percentage transfer whether it takes 50 , 100 , 500, 1000 hours, or whatever. To make the formula sensitive to time invested in the experimental condition, Roscoe introduced the cumulative transfer effectiveness ratio (CTER)

$$\text{CTER} = \frac{Y_c - Y_x}{X}$$

Y_c and Y_x are defined as before. X is the time, trials, or errors required by an experimental group during the experimental treatment (e.g., simulator training). This formula is sensitive to the cost of time invested in the experimental condition; for example, if no gain is made in the time required to reach criterion for 50, 100, 500, or 1000 hours on the experimental task, the CTER will drop from 4 to 2 to 0.4 to 0.2. Of course, this would be an extremely unlikely occurrence. There are certain general expectations about how differing amounts of simulator training will affect later performance on the actual task; Roscoe states:

> There is convincing inferential evidence that successive pre-solo hours in a ground trainer yield decreasing increments of saving in pre-solo flight time, and the same decreasing incremental benefits would be expected for any successively related educational experience (p. 4).

In other words, there is a better return for early hours of simulator training than for later hours.

The CTER is not particularly useful when taken as a snapshot with a single set of values. The most effective way to apply it in CEAT is to estimate CTER for a range of values of X and the associated costs. One of the first applications of CTER is in Povenmire and Roscoe (1972). Figure 7 is adapted from data they reported for an experiment concerning transfer from a flight simulator to flight training. The figure shows the typical form of the CTER function.

---

[4] Roscoe put it this way: "For reasons beyond comprehension, there has been no recognition in the psychology of learning of the intuitively obvious fact that the effectiveness of transfer is also a...function of the amount of such practice" (p. 3).

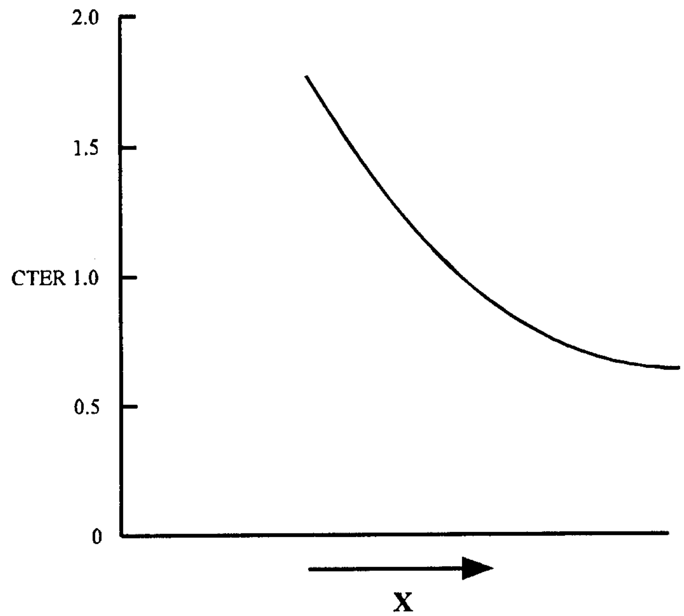


Figure 7. **Typical form of CTER function when plotted over several values of X. (Adapted from Povenmire & Roscoe, 1972.)**

There are criticisms of CTER. One of the concerns is that ratios can be difficult to comprehend intuitively and can be misleading (DoD, 1991b). The general guidance is that if ratios are used during CEAT, raw data should also be presented so that they can be interpreted separately. Boldovici (1987, 1993) contends that ratios have several problems. Among other things, a given ratio (e.g., CTER = 1.0, indicating equal effectiveness for two forms of training) can be produced in several different ways. He offers the example of a control group taking 20 trials to reach criterion and an experimental group taking 12 trails on the simulator and 8 on the equipment, yielding a CTER of 1.0. A second experimental group taking 2 trials on the simulator and 18 on the equipment, would also yield a CTER of 1.0. The training effectiveness of both experimental treatments is identical according to the CTER, but the conditions are obviously different. Boldovici also offers several criticisms on methodological grounds.

CEAT Conceptual Model

Figure 8 shows how the CEAT concepts described in this section fit together into a CEAT conceptual model. This is a graphic representation of the sequence of steps described earlier for CEAT with additional elements included and arrows showing process flow. The first step in the process is to formulate assumptions regarding what variables will affect the process and the range of values those variables will present.

The next step is to determine what training alternatives will be compared (e.g., new system and one or more other possibilities). (The formal CEAT definition calls for alternatives to be compared, but in practice this is not always done; a CEAT that does not compare alternatives is compromised.)

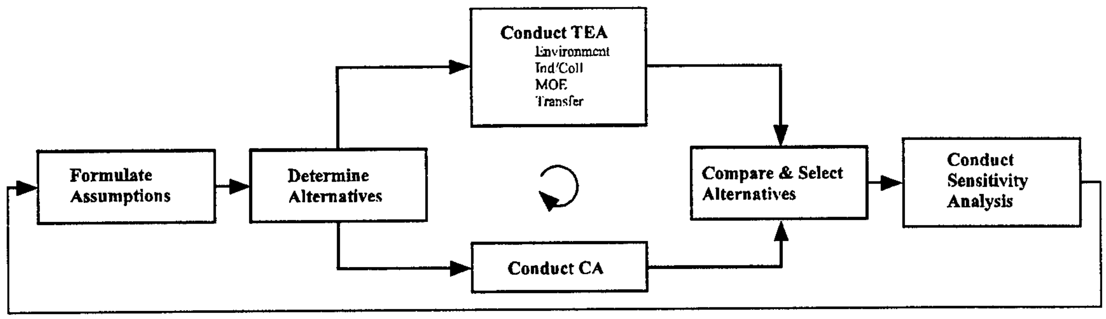


Figure 8. **CEAT conceptual model showing sequence of steps and process flow in idealized CEAT.**

Next, TEA and cost analyses (CA) are performed. These steps occur more or less concurrently, as the CA is performed for the same training alternatives considered in the CA and there will be some interplay between the two analyses. TEA occurs within a particular training environment (e.g., school or unit), with particular types of tasks (e.g., individual and/or collective), whose performance is measured using MOE. Ideally, the TEA uses MOEs that measure transfer to the operational setting. One of the essential steps in the TEA is to use a formal method to assign values to MOEs. (Several different methods are used, as discussed in the *CEAT Methods* section of this report.) The foregoing is a highly simplified description of TEA but illustrates the general logic.

CA are performed to determine the costs of the alternatives. This can be done in several different ways, as discussed in *CEAT Methods*.

The alternatives are then compared and the best alternative is selected. This is done using formal decision rules such as Orlansky's decision logic (Figure 1). This step is followed by sensitivity analysis. Sensitivity analysis modifies the assumptions and recycles the CEAT process. This may occur iteratively, several times, as the process is tuned to find a training alternative that truly is the "best." The process then ends. (As in other departures from the ideal, sensitivity analysis is not always performed.)

The next section, *CEAT Methods*, discusses TEA, CA, and sensitivity analysis in greater detail.

CEAT METHODS

The *CEAT Concepts* section described several CEAT concepts and a general CEAT method. The present section describes CEAT methods in greater detail. The first subsection discusses the impact of time on CEAT. The second subsection presents a taxonomy of TEA methods which illustrates the many different ways that TEA can be conducted. Subsections following describe several different empirical and analytical TEA methods as they relate to the taxonomy. The following two subsections discuss cost analysis and sensitivity analysis. The final subsection discusses the dilemma faced in choosing among CEAT methods.

Impact of Time on CEAT

Time affects CEAT significantly. If a training system has not been built, it cannot generate empirical training data. It has been estimated that approximately 75% of a system's acquisition cost has been committed by phase II of the development cycle (Zimmerman, Butler, Gray, Rosenberg, & Risser, 1984). Concurrently with the expenditure of funds, two other things are happening: availability of training data is increasing and the potential for change is decreasing (Klein, Johns, Perez, & Mirabella, 1985). Notional relationships among expenditures, availability of training data, and potential for change of a training system during the WSAP are illustrated in Figure 9. There is an incentive to conduct CEAT early to save funds and to identify needed changes as early as possible.

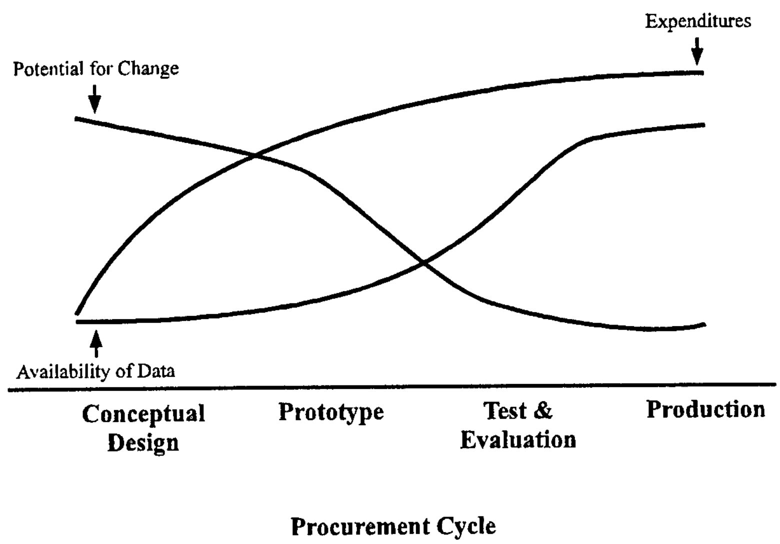


Figure 9. **Notional relationships among expenditures, availability of training data, and potential for change of training system during the WSAP. (Adapted from Klein et al., 1985.)**

Matlick, Berger, Knerr, and Chiorini (1980) identified six different data input situations which are likely to obtain as the WSAP progresses:

- No task list and no training program
- Task list but no training program
- Training program but no alternatives and no effectiveness data
- Training program with effectiveness data but no alternatives
- Alternative training programs but no effectiveness data for all alternatives
- Training program alternatives and effectiveness data for all alternatives

The relationships among these steps are illustrated in Figure 10. This analysis has been used elsewhere (e.g., Matlick, Berger, & Rosen, 1980; Rosen, Berger, & Matlick, 1985; Knerr, Nadler, & Dowell, 1984) and remains a useful framework for considering how time affects available data and the analytical methods possible in a CEAT. The first question Figure 10 poses is whether or not a task list exists. If not, one must be generated. The next question is whether or not a training program (or device or system) exists. Regardless of the answer, training alternatives are then considered; this may require that they be identified. The availability of training effectiveness data for the alternatives is considered, then cost and cost-effectiveness.

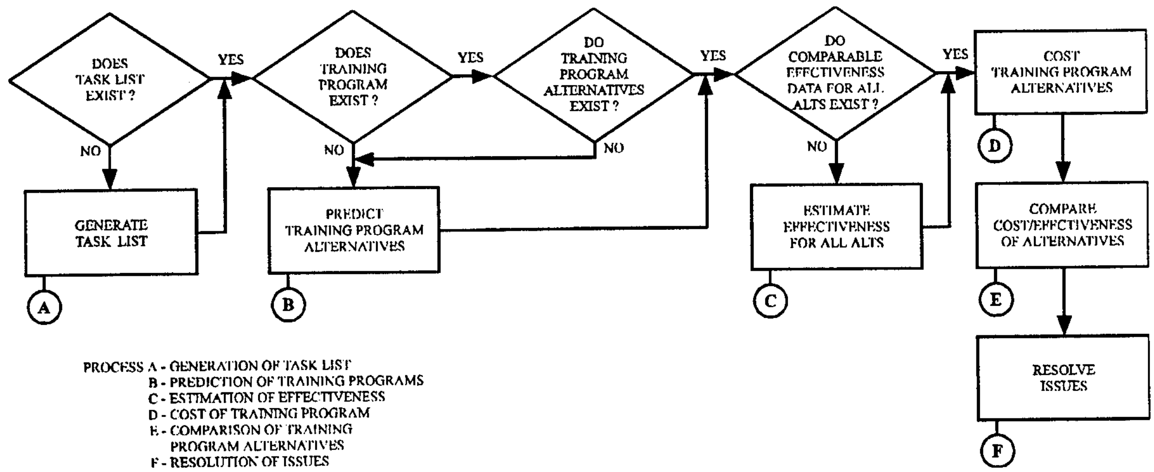


Figure 10. **A general CEAT model. (From Matlick, Berger, Knerr, & Chiorini, 1980.)**

Often the answers to some of the questions Figure 10 poses will be negative. Consider that the only way it would be possible to perform an experimental comparison among alternatives would be if training tasks, programs, and alternatives existed. If a task list existed but no alternatives, one would either have to create an alternative and collect data or contrive some way to estimate training effectiveness data (e.g., have SMEs make estimates, look at similar systems). (A less desirable option would be to assess the new system based on how well it met its objectives, without comparing it to an alternative, though this runs counter to the definition of CEAT in *CEAT Methods*.)

If there is no task list, it is even more difficult to perform the analysis; everything must be estimated.

The *CEAT Methods* section made the point that transfer of training is important in TEA. Martin, Rose, and Wheaton (1988) observe that empirical transfer of training studies are recognized as the traditional method of assessing the effectiveness of training devices but that such studies frequently cannot be used; the alternative is to use analytic methods to predict transfer. (Another less desirable alternative is to assess training effectiveness without attempting to determine transfer.)

Finally, consider that the availability of cost data depends upon everything else being in place. It is difficult to estimate the cost of a training program, system, or device that does not exist.[5] The problem is analogous to that for TEA. Empirical comparisons can be made if concrete examples of the alternatives exist; otherwise, some part of the TEA must be based on estimates.

[5] Rankin (1994) notes that it is possible to make estimates using work breakdown structures.

The equivalent in cost estimation to training programs in TEA is the existence of historical data.[6] The availability of historical data for two or more alternatives is a rare occurrence. When the data are not available, they must be estimated.

TEA Methods Taxonomy

The *CEAT Methods* section described MOE as tools to assist in discriminating among training alternatives but did not tell how value would be assigned to MOE during the TEA. Historically, four different methods have commonly been used:

- Conduct an experiment
- Estimate from data based on similar systems
- Obtain SME estimates
- Apply an analytical method

The first three methods rely on observation and/or experience and hence formally are empirical methods. They vary in credibility. Experiment has the greatest face validity. The relative persuasiveness of data based on similar systems vs. SME estimates depends upon the situation. In general, the cost of these methods is highest for experiments and lowest for SME estimates, with estimates based on similar systems in the middle. (Empirical methods are described in greater detail later.)

Analytical methods are a class of non-empirical methods designed to do many different things. One of these things is to model a training system such that its behavior under different input conditions can be predicted. In CEAT, the idea is to use the model to determine the effect of input conditions on MOE. (Analytical methods are described in greater detail later.)

The distinctions among the non-experimental methods are not necessarily obvious, and depend somewhat upon definitions and interpretations. For example, analytical methods usually are driven by empirical data (e.g., SME estimates, ratings, etc.) However, typical analytical methods include elaborate problem definition and analysis phases that structure the problem and process the data in a more complex manner than would be done with the other non-experimental methods. They usually employ algorithms, decision rules, and mathematical formulas (Goldberg & Khattri, 1987). Methods which use data from similar systems are sometimes classified as analytical methods. However, it is useful to classify them separately because the other analytical methods do not usually use comparison systems. (In some cases, comparisons may be performed, but comparison is not the main operating mechanism of the method.)

[6] The use of historical cost data to estimate cost in a new situation is known as "costing by analogy." Cost analysis is discussed in greater detail later.

Given that there is a method to assign values to MOE, a reasonable question to ask is what are the MOE values going to be used for? Jeantheau (1971), as cited in Knerr et al., distinguishes among these four levels of evaluation:

- Qualitative
- Non-comparative
- Comparative
- Transfer

Implicit in the distinctions is that higher levels (e.g., transfer) are more authoritative than lower levels (e.g., qualitative). Qualitative evaluation is typically based on subjective estimates which do not assign quantitative value. For example, one might rank a training system attribute as "good" but be unable to say how good in any absolute sense.

Non-comparative evaluation assigns value based on a set of standards. This is commonly done in the world of training development. Quantitative value can be assigned. An example would be to conduct a pilot course and evaluate its effectiveness based on the percentage of training objectives met to standard by students.

Comparative evaluation assigns value to two or more competing training alternatives. Quantitative value can be assigned. At the end, the values obtained enable one to pick the winner.

Transfer evaluation assigns value based on performance in a new situation. An example given in the *CEAT Methods* section was transfer from a flight simulator to in-flight performance. If two alternatives are being compared, the winner is the one with the greatest percentage of transfer.

An actual evaluation may involve some combination of these levels. Note that the formal definition of CEAT given in the *CEAT Methods* section calls for the comparison of alternatives. To the degree that one abides by the formal definition, it would seem that a CEAT should include comparative and transfer evaluations. Qualitative and non-comparative evaluations are not necessarily ruled out, but should not be the primary means by which CEAT data are obtained.

Note that the methods and levels of evaluation cross to form a matrix that is a useful framework for conceptualizing how TEA might be performed (Table 5). Conceptually, the different evaluation methods can be used to obtain different levels of evaluation data. Some cells represent more likely method-level combinations than others (shown in bold). For example, the C-E (comparative-experimental) combination may be the one most people think of in relation to TEA; provide training treatments A and B under experimental conditions and select the winner based on MOE scores. The qualitative-SME (Q-S) combination is used every time one relies exclusively on the judgment of an SME to assess training effectiveness; this happens virtually daily in military schools. The NC-C combination represents a situation in which one predicts how well a new training system will meet training standards based on its similarity to an existing system. Formally or otherwise, and consciously or not, evaluators rely heavily on their experience with existing systems in evaluating new ones. Analytical methods which can predict transfer (T-A combination) would be extremely useful and much effort has been spent building them.

Table 5

TEA Framework Relating Evaluation Methods and Levels. Conceptually, the different evaluation methods can be used to obtain the different levels of evaluation data.

Methods		Levels			
		Qualitative	**Non-Comparative**	**Comparative**	**Transfer**
Empirical	Experimental	Q-E	NC-E	C-E	T-E
	Comparison	Q-C	NC-C	C-C	T-C
	SME	Q-S	NC-S	C-S	T-S
Analytical		Q-A	NC-A	C-A	T-A

Some combinations strain conventions. The non-comparative-experimental combination (NC-E) implies conducting an experiment with one condition. This may satisfy the dictionary definition of experiment ("test") but not the usual scientific requirement of having a control as well as an experimental group.[7]

Other combinations make sense logically, though they require one to think about TEA in somewhat novel ways. Examples are using a comparison system to predict transfer to a new system (T-C combination), using an analytical method to predict qualitative data (Q-A combination), etc.

Empirical Methods

Experiment

Hoffman and Morrison (1992) and Pfeiffer and Browning (1984) provide excellent overviews of several alternative experimental designs that have been used for TEA; this subsection is based mainly on their explanations. Experiments, if well conducted, provide the most persuasive evidence of training effectiveness (Morrison & Hoffman, 1992). In addition, they may provide strong evidence to justify budgets; comply with acquisition, test, and evaluation regulations; and may lead to training improvements (Boldovici & Bessemer, 1994).

However, it is almost always challenging to conduct experiments well in an operational setting. Among the more obvious problems are high cost, lack of experimental control, and difficulty of manipulating events for experimental purposes (Hoffman & Morrison). These factors

[7] "Experiments" with one condition are known as "quasi-experiments" (or demonstrations). They are sometimes used when the situation seems to preclude traditional testing methods (e.g., to test one-of-kind systems for which it is complex, expensive, or dangerous to obtain comparison data.)

may compromise experiments and threaten valid inferences because of too few subjects, differences among groups being compared, confounding treatments, and other factors (Boldovici, 1987). Bessemer (1991) contends that many of the factors that confound comparisons in experimental research (as identified in Campbell & Stanley, 1966; and Cook & Campbell, 1979) threaten operational research relating to SIMNET (e.g., history, maturation, instrumentation, selection, mortality, causal direction). By reasonable extension, these factors threaten experiments in operational settings more generally. Full discussion of these problems is beyond the scope of this report; readers should refer to the cited works for additional information.

The problems in conducting research in operational settings have led researchers to use ingenuity in their designs as well as to consider non-experimental methods as alternatives. Experimental designs are sketched below as they relate to the categories in Table 5. The designs are represented graphically in Table 6.

Table 6

Graphic Representations of TEA Designs. (Adapted from Pfeiffer & Browning, 1984.)

Example	Experimental levels	Group(s)	Design
a	Non-Comparative	Experimental	Simulator----->op. equipment
b	Non-Comparative	Experimental	Pretest-->simulator-->postest
c	Comparative	Control & experimental	Experimental training--->testing Control training--->testing
d	Comparative	Control & experimental (several groups)	Experimental training #1 --->testing Experimental training #2 --->testing Etc. (additional experimental groups) Control training--->testing
e	Transfer (Snapshot)	Control & experimental	SIM----->A/C --->A/C
f	Transfer (Function)	Control & experimental (several groups on the same simulator)	SIM->A/C SIM--->A/C Etc. (additional SIM groups) --------->A/C
g	Transfer (several SIM conditions)	Control & experimental (several groups on different simulators)	SIM#1->A/C SIM #2--->A/C Etc. (additional SIM groups) --------->A/C

Qualitative Experimental (Q-E). Q-E design is used to obtain data in many situations. Perhaps the most common is the rating form used by students to evaluate different attributes of a course at its conclusion. Data thus obtained can be used to assess training against a standard, compare training alternatives, etc.

Non-Comparative-Experimental (NC-E). NC-E designs are usually referred to as "quasi-experimental," where "quasi" suggests resemblance to experiment without meeting all the usual requirements.[8] Pfeiffer and Browning give examples of designs in which the performance of a single group is tested at two separate points (e.g., first on simulator, later on equipment; using pre- and post-test on simulator), in which training effect is inferred if performance improves from first testing to second (Table 6a, b). Hoffman and Morrison refer to the latter as a "two-point assessment", useful for determining performance improvement on a device. Boldovici and Bessemer endorse the use of such designs, preferably in conjunction with others, "to provide converging evidence compensating for various weaknesses of each method used alone" (p. 23). Such designs are less expensive alternatives to more traditional (and what the authors regard as frequently flawed) designs. Inferences can be drawn with such designs, but they are weaker than when a traditional control group is used.

Comparative-Experimental (C-E). C-E designs compare one or more experimental groups with a control group (Table 6c, 6d). The designs compare performance among conditions but does not assess transfer to the ultimate operational setting (e.g., job performance, equipment operation). Strong inferences can be drawn from these designs, within their limits, and they are less complex than transfer designs.

Transfer-Experimental (T-E). Transfer of training was characterized in the *CEAT Methods* section as the degree to which training in situation A prepares one to perform in situation B. Situation A typically involves formal training (e.g., with a simulator) and situation B is usually the operational setting (e.g., with operational equipment), though the basic paradigm applies between any two situations. (The following discussion refers to simulators (SIM) and aircraft (A/C) to make it simple and concrete but is meant to apply more generally.) Several different T-E designs are possible. The simplest takes a single snapshot look at transfer. One group receives training on SIM and the other on A/C. The SIM group eventually moves to A/C (Table 6e). Both groups are tested on A/C. If the SIM group takes less time on the A/C than the A/C group to reach a criterion level of performance, positive transfer has occurred. A positive finding validates occurrence of transfer but does not indicate the functional relationship between amount of SIM training and amount of transfer. To do this, several different SIM conditions, representing different amounts of training time on the same simulator, are required (Table 6f). Another variant of T-E design compares transfer on several different simulators (Table 6g).[9]

---

[8] According to Cook and Campbell, the term is used in reference to experiments that have treatments, outcome measures, and experimental units, but that lack random assignment. Comparisons are made on nonequivalent groups and the researcher has to explicate factors which threaten valid causal inference and deal with them in some reasonable way.

[9] Pfeiffer and Browning also sketch other transfer designs for use in TEA; the interested reader should refer to the source for additional information on the subject.

Comparison-Based Methods

Comparison-based methods estimate the training effectiveness or cost of a new system based on its similarity to and differences from comparable systems, adjusting upward for positive attributes of the new system and downward for negative attributes in relation to the comparison system. The Hardware/Manpower Integration Program (HARDMAN) uses a comparison-based method to predict manpower, personnel, and training requirements for a new system based in part on its similarity to a baseline system (Department of the Navy, 1987). An important comparison-based method in CEAT is Klein Associates' Comparison-Based Prediction, which Klein et al. describe as follows:

> Comparison-Based Prediction (CBP) is a method of reasoning by analogy, where an inference is made for one object or event based upon a similar object or event. It is the use of concrete experience as a basis for predicting the future, making adjustments on the basis of key differences between the cases (p. 1-4).

CBP can be used to predict training effectiveness, cost, or both. Klein et al.'s description of the procedure for applying this method suggests that it could be used to obtain empirical data at any of the four levels (qualitative, non-comparative, comparative, transfer) for it makes predictions based on SME opinions in a structured data collection process. The examples given, however, are non-comparative (NC-C) (i.e., one prediction is made for one hypothetical system). However, there is no reason CBP could not be used to make predictions for several different systems and then to compare the outcomes.

Elements of the CBP methodology are illustrated in Table 7 as they relate to the home appraisal process.

The steps in CBP (from Klein et al.) are as follows:

1. Specify the target (A), the device whose training effectiveness or cost is to be predicted.
2. Define the measure (T) of training effectiveness or cost to be predicted.
3. Identify the major causal factors (high drivers) that will affect the target variable for A, T(A).
4. Determine the conditions under which A will operate.
5. Identify device(s) (B) which will be used for comparison.
6. Select a CBP strategy.
7. Select SMEs.
8. Determine, with SMEs, the comparison value, T(B), for comparison device(s)
9. Examine scenario differences between cases A and B. Estimate effects of differences on T(B).
10. Adjust value of T(B) to allow for differences between B and A.
11. Determine a value for T(A) from this adjustment.
12. Document the process to provide an audit trail.

Table 7

Elements of Comparison-Based Prediction Methodology. (Adapted from Klein et al., 1985.)

Home appraisal element	CBP element
Home being sold	Target case: A
Selling price	Target variable: B
Selling price for A	Target value: T(A)
Appraiser	SMEs
Other comparable homes, previously sold	Comparison case(s): B
Factors that may influence selling price of A (e.g., size, age, number of rooms)	Causal factors (from which high drivers are selected)
Final list of most important factors, their specific values, and how they affect one another	Scenario
Decision on how many comparison houses (B) to use and how many and what kinds of appraisers to use	Strategy
Selling price for a comparison house	Comparison value: T(B)
Documentation/Report on how selling price of target house was estimated	Audit trail

As is apparent, CBP relies on SME judgments and could reasonably be characterized as an SME-based method. It is more formally structured than most of the variants of SME methods and hence is classified separately here.

Klein et al. indicate that CBP has been used in several different applications and at least partially validated. Among its strengths are its applicability early in the WSAP and relative low cost. Its limitation is reliance on comparison systems; if none are available, the method is inapplicable (Pfeiffer & Horey, 1988). Goldberg (1988) reviews CBP as follows:

> In our opinion, CBP has identified an area worth considering and formalized a process for doing so: the situation in which there is a similar TD/S [training device/simulator] from which estimates can be made for a newly developing TD/S. However, a great deal of the process as described represents defining the problem, much like any other problem solving process (p. 31).

Notably, Goldberg, Adams, and Rayhawk incorporated elements of CBP in the Training Effectiveness and Cost Iterative Technique (TECIT), a CEAT method they developed for the ARI.

SME-Based Methods

SME-based methods, for purposes of this report, are methods that rely primarily on SME estimates, ratings, or other indicators to provide data to assess training effectiveness. SMEs can provide data in several different ways. For example, in a comparison of two different training methods, probably the simplest way would be to have SMEs observe the training treatments and then express their opinion on which was "best." Such a declaration would be of questionable value as it leaves it up to SMEs to choose the decision criteria and provides only categorical data. The decision process can be improved by structuring it to define decision criteria, collecting data in a manner that allows ordinal data to be collected (e.g., with checklists, rating forms, or other methods to scale the data), and leaving it to an analyst to decide what "best" means according to formal decision rules which are applied later.

Pfeiffer and Horey (1988) identify two methods that fit this report's definition of SME-based methods: checklist and instructional quality inventory (IQI). The authors categorize them as "index" methods because they scale data by counting the number of attributes present and hence yield ordinal data.

A checklist can be used to evaluate training systems or devices. It consists of a list of statements describing desirable attributes. The list is presumably compiled by experts who base it on research findings, empirical data, historical precedent, widely-respected design standards, or some combination of these factors. SMEs observe training, consider each item on the list, and decide whether to check the item "yes," "no," or "n/a." At the end, the number of "yes"'s on the checklist is tallied to obtain a score. The higher the score, the better.

A variation of the checklist is to rate training attributes on a scale (e.g., rate the quality of training aids used in a course on a scale from 1 to 10). Combining the scores of rated items is more complicated than tallying "yes"'s but can also yield interval data.

Checklists, ratings, and other scaling methods can be used to obtain data at the first three levels in Table 5 (qualitative, non-comparative, comparative). In principle, SMEs could also estimate transfer.

The IQI is designed to assess formal schoolhouse training courses. It is a subjective questionnaire for assessing course learning objectives, test items, and instructional materials. The method provides checklists and rules to assess adequacy of each element and desirable relationships among elements. Procedures are provided to determine adequacy of objectives in terms of content and instructional intent; test item adequacy in terms of consistency with corresponding objectives and test item construction; and presentation consistency and adequacy.[10] IQI was designed for use in a non-comparative manner; that is, to assess training against a set of standards (NC-S method-level combination in Table 5). Nonetheless, it could be used to compare two different training courses if training materials for both were available.

Frederickson (1981) describes a complete CEAT method developed by Applied Science Associates that is driven by SME data. The method consists of 10 tasks:

1. Prepare work plan
2. Analyze missions and functions
3. Select tasks for training
4. Analyze tasks
5. Generate general course structure
6. Generate training program alternatives
7. Develop extended program of instruction
8. Analyze training effectiveness and trainability
9. Analyze training costs
10. Conduct final tradeoff analysis

During step 4, tasks are ordered in terms of 10 criticality dimensions (e.g., time delay tolerance, consequences of inadequate performance, immediacy of performance, importance, frequency of performance). The first seven steps are primarily front-end analysis and yield, among other things, a set of tasks to be covered in training, their criticality dimensions, and alternative programs of instruction (POI). At step 8, SME's estimate the effectiveness of training each task for each POI alternative, where effectiveness is defined as the percentage of students that would reach the performance criterion for each time condition. SME estimates are used to assess the trainability of each task. The method then takes into account the criticality dimensions for each task to develop a figure of merit for the alternative POI. Five of the 10 criticality dimensions are applied, using the scale values shown in Table 8. According to Frederickson, the five dimensions were scaled based on the "information utility they provide for determining the worth of including a task in a training program" (p. 437). He does not explain exactly how these numbers were derived.

This method, as described, provides comparative data (C-S method-level combination in Table 5). It could be used non-comparatively (NC-S). It does not estimate transfer to the operational setting.

[10] IQI later evolved into the Course Evaluation System (CES). In 1988, CES was used to evaluate 100 Navy courses with the following interesting results, as reported in Taylor, Ellis, and Baldwin (1988): "56 percent of the 1945 knowledge objectives examined were inappropriate for the course training goal and future job requirements...49 percent of the objectives were not tested...48 percent of all test items did not match related objectives...38 percent of all test items were inappropriate...practice was incomplete or not present for almost one-half of the presentations...many instructional strategies proven to be effective in civilian classrooms were not utilized" (p. iii).

Table 8

Task Criticality Dimensions and Scale Values Used in Applied Science Associates' SME-Based CEAT Methodology. (From Frederickson, 1981.)

Criticality dimension	Scale value
Consequences of inadequate performance	0.45
Task importance	0.26
Time delay tolerance	0.16
Frequency of performance	0.08
Immediacy of performance	0.05

Analytical Methods

Authors of a review of the analytical CEAT literature once declared, in a tone of apparent exasperation, that "the proliferation of models and methods...has created a body of work that is extensive and bewildering. Analysts charged with the conduct of [analyses] require systematic classification and evaluation of the methods" (Rosen et al., 1985, p. 2-44). The present study's literature review validated Rosen et al.'s impression by revealing dozens of different yet often linked or related methods that evolved across time. In 1994, Muckler and Finley published a two-volume review that describes and compares 36 of the methods clearly and concisely from a historical perspective for the decade 1970-1990; this review goes a long way toward sorting out the field and is recommended to readers interested in its historical development.[11]

Because of the complexity of the picture, this subsection will discuss it historically rather than in terms of method-level combinations. The methods are described here as they were revealed in the literature for the decade from 1980-1990. The story can be traced earlier than 1980, but to do so would add detail without necessarily clarifying the current state of analytical methods in the DoD. Events since 1990 are discussed briefly in the next subsection. The following summary does not purport to be comprehensive but to sketch the main developments in analytical methods during the decade covered.

The distinction is commonly made between predictive and prescriptive analytical methods (Knerr, Nadler, & Dowell, 1984; Goldberg & Khattri, 1987; Martin & Rose, 1988). Prescriptive models tell how training should be conducted, while predictive models predict training effectiveness given that training is conducted in a particular way. In TEA, the intent is to evaluate training systems before procurement; that is, while they are prototypes or purely conceptual in nature; hence, the interest is in predictive methods. All of the analytical methods discussed in this subsection are predictive, and some (such as OSBATS and TECIT) are also prescriptive.

[11] Volume I (Muckler & Finley, 1994a) contains a literature review and analysis and volume II (1994b) contains a 175-item annotated bibliography that covers the essential literature in the field. The author is indebted to Jesse Orlansky for bringing the review to his attention.

Historical Perspective

Figure 11 illustrates the most influential analytical method reviews and methods and their apparent relationships as revealed in the literature review. Report authorship shows that many of the principals involved in method development tend to stay with a particular method across time. Examples are Sticha with OSBATS (optimization of simulation-based training systems); Rose and Martin with DEFT (device effectiveness forecasting technique); Matlick, Berger, and Knerr with a series of reviews in the early 1980s and an early method that later evolved into TECIT under the guidance of Goldberg and others. TRAINVICE evolved into DEFT, which later evolved into ASTAR. Goldberg (1985) acknowledges that TECIT incorporates elements of DEFT, FORTE, and CBP. There have been two projects within the Army to develop methods that do not fit conveniently within the framework of Figure 11; these are discussed later in this section under the heading "Other Developments."

A MATRIS summary of work units and studies and analyses in CEAT (Smith, 1994) revealed that Sticha had last been contracted to perform major work on OSBATS in 1991 and that none of the other principals was shown to be under contract with the government for work in the area. Sticha continues to refine OSBATS on a limited scale for the ARI (Sticha, 1994). Pfeiffer, the architect of FORTE, is deceased, but FORTE is still used occasionally by personnel at the Naval Air Warfare Center, Training System Division (NAWCTSD) (Micheli, 1994). ASTAR (automated simulator test and assessment routine), which evolved from DEFT, was evaluated in 1990 and declared unready for widespread implementation (Companion, 1990). For many years the ARI has supported research relating to analytical methods. The scope of research has been reduced recently as declining resources have forced the Army to focus on work deemed by the Service to be of higher priority (Singer, 1994).

The reviews shown in Figure 11 focused on various existing methods and models. Nearly 50 different methods and models were reviewed. With a few exceptions, most are now mainly of historical interest. The listed terms are defined in *Abbreviations and Acronyms*, but are not discussed in this report. Readers interested in further information should consult the reviews themselves.

OSBATS

Sticha and colleagues developed OSBATS under contract to the Army as a computer-based tool to help designers conduct tradeoff analyses to produce "cost-efficient" training devices (Sticha, Blacksten, Buede, and Cross, 1986; Sticha, Blacksten, Knerr, Morrison, and Cross, 1986). OSBATS includes models to structure the design problem, specify the decision process, define data content and format (normative models); and models to predict performance (descriptive models) (Sticha, 1989). Sticha states that "the overall modeling framework is based on methods that attempt to define the training strategy that meets the training requirements at the minimum cost...in its simplest form, the method compares the ratio of effectiveness of two training alternatives to the ratio of the cost of the options" (p. 457). Sticha credits Roscoe (1971) with originating the framework, and others who extended it, before OSBATS was developed.

OSBATS is driven by data about training requirements, task characteristics, trainee population skills, training device instructional features, and fidelity dimensions. Much of this data must be estimated by SMEs (Sticha, 1989).

Sticha, Singer, Blacksten, Morrison, and Cross (1990) describe OSBATS as consisting of five modules:

> 1. Simulation Configuration Module. A tool that clusters tasks into the categories of part-mission training devices, full-mission simulators, and actual equipment.
> 2. Instructional Feature Selection Module. A tool that analyzes the instructional features needed for a task cluster and specifies the optimal order for selection of instructional features.
> 3. Fidelity Optimization Module. A tool that analyzes the set of fidelity dimensions and levels for a task cluster and specifies the optimal order for incorporation of advanced levels of these dimensions.
> 4. Training Device Selection Module. A tool that aids in determining the most efficient family of training devices for the entire task group, given the training device fidelity and instructional feature specifications developed in the previous Modules.
> 5. Resource Allocation Module. A tool that aids in determining the optimal allocation of training time and number of training devices needed in the recommended family of training devices. (p. 15.)

In application, these modules are intended to be used iteratively to arrive at an optimum solution to the design problem. Sticha provides this description of an analyst using OSBATS to decide how training should be conducted:

> ...the analyst uses the Simulation Configuration Module to examine the tasks to be trained and to provide a preliminary recommendation for the use of either actual equipment or one or more training devices. The result of this analysis is three clusters of tasks. Two of these clusters define tasks for which a full-mission simulator or part-mission training device should be designed.... The analyst then uses the clusters [in] the Instructional Feature Selection and Fidelity Optimization Modules [which] define...a range of options that vary in cost.... The Training Device Selection Module evaluates...device design[s].... The analyst exercises this module several times using different combinations of training devices.... [When satisfied, the analyst] investigates the solution using the Resource Allocation Module.... (p. 16)

Although OSBATS is intended as a design tool, it can be used to compare hypothetical designs (the Analytical-Comparative method-level combination in Table 5). Its algorithms predict transfer from device to operational equipment. Hence, it also meets the requirements for

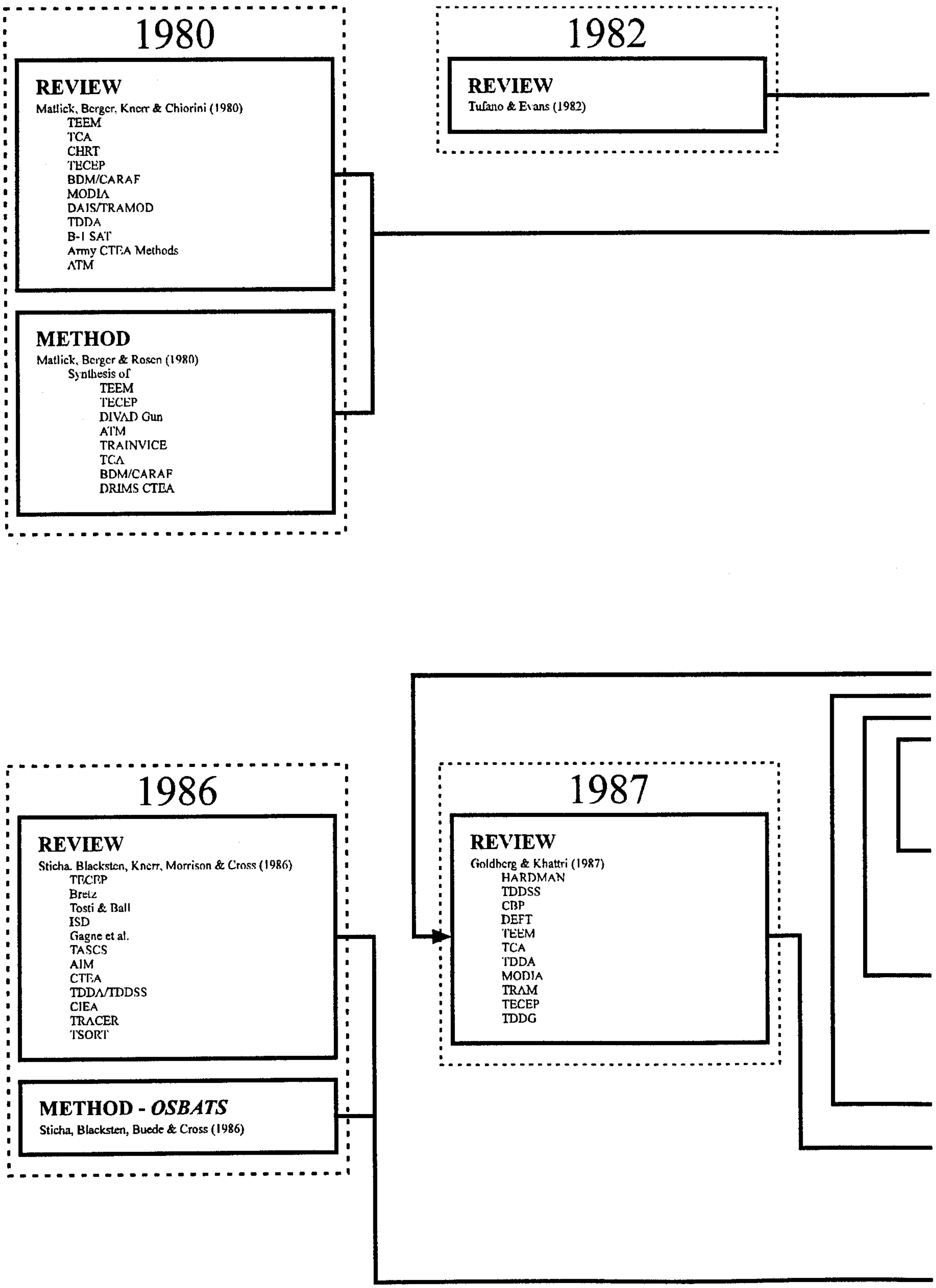
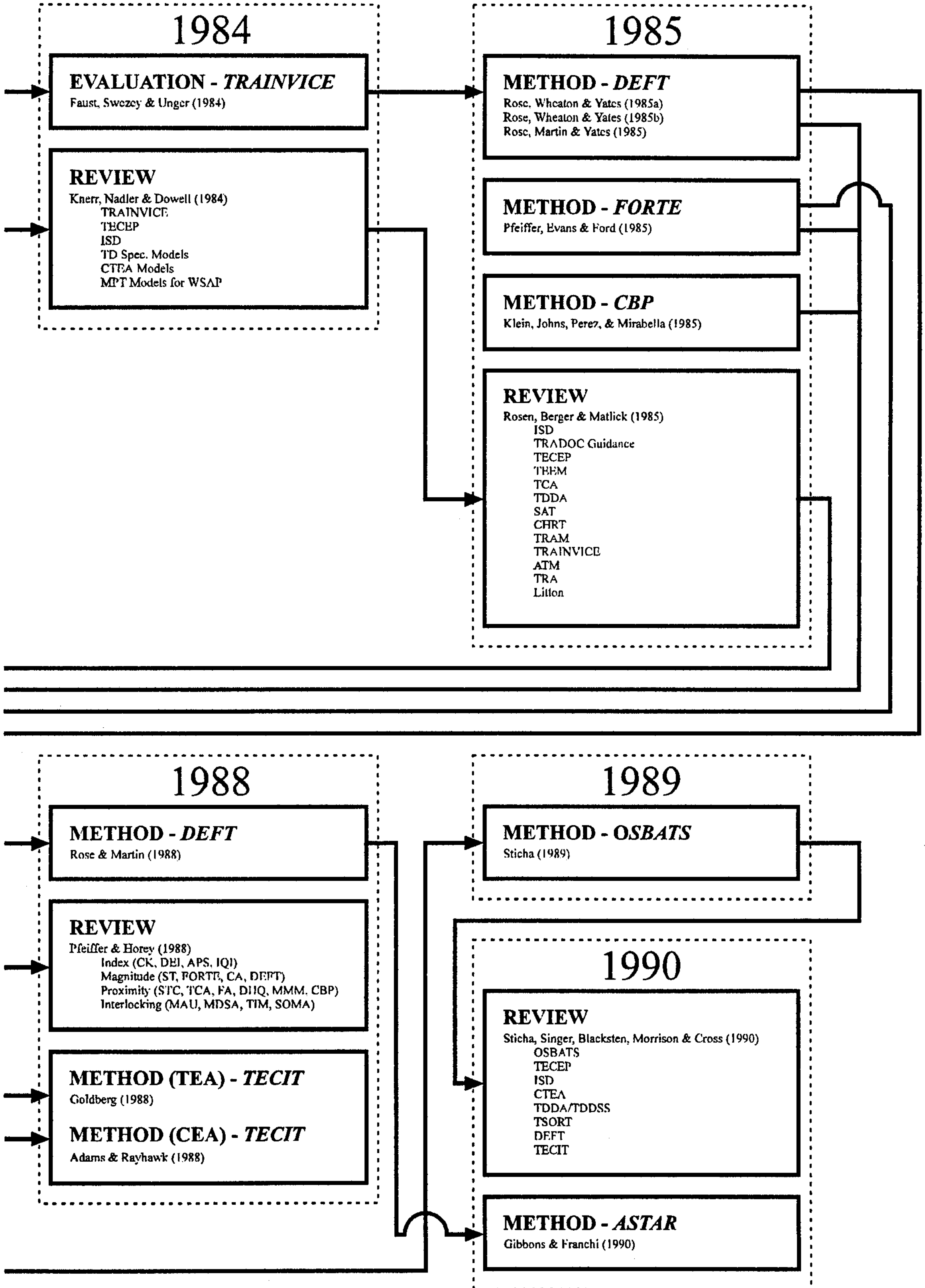


Figure 11. **Historical perspective on analytical methods used in CEAT.**


1984
EVALUATION - *TRAINVICE*
Faust, Swezey & Unger (1984)
REVIEW
Knerr, Nadler & Dowell (1984)
TRAINVICE
TECEP
ISD
TD Spec. Models
CTEA Models
MPT Models for WSAP
1985
METHOD - *DEFT*
Rose, Wheaton & Yates (1985a)
Rose, Wheaton & Yates (1985b)
Rose, Martin & Yates (1985)
METHOD - *FORTE*
Pfeiffer, Evans & Ford (1985)
METHOD - *CBP*
Klein, Johns, Perez, & Mirabella (1985)
REVIEW
Rosen, Berger & Matlick (1985)
ISD
TRADOC Guidance
TECEP
TEEM
TCA
TDDA
SAT
CHRT
TRAM
TRAINVICE
ATM
TRA
Litton
1988
METHOD - *DEFT*
Rose & Martin (1988)
REVIEW
Pfeiffer & Horey (1988)
Index (CK, DEI, APS, IQI)
Magnitude (ST, FORTE, CA, DEFT)
Proximity (STC, TCA, FA, DIIQ, MMM, CBP)
Interlocking (MAU, MDSA, TIM, SOMA)
METHOD (TEA) - *TECIT*
Goldberg (1988)
METHOD (CEA) - *TECIT*
Adams & Rayhawk (1988)
1989
METHOD - *OSBATS*
Sticha (1989)
1990
REVIEW
Sticha, Singer, Blacksten, Morrison & Cross (1990)
OSBATS
TECEP
ISD
CTEA
TDDA/TDDSS
TSORT
DEFT
TECIT
METHOD - *ASTAR*
Gibbons & Franchi (1990)

the Analytical-Transfer method-level combination. The method focuses on training devices, particularly those applicable to aviation training. It is unclear how applicable it is in the optimization of other forms of training.

Sticha (1989) commented on the difficulty of validating OSBATS as follows: "Because of the complexity of the OSBATS model, validation of the model as a whole is probably impossible. Other aspects of the model preclude validation of major sections of the model without empirical data...Probably a better strategy is validation of submodels to determine key model parameters" (p. 465).

OSBATS has not been validated, although a formative evaluation was reported in 1990 (Sticha, Blacksten, Buede, Singer, Gilligan, Mumaw, & Morrison); this evaluation led to several suggestions for improvements and further tests. Singer (1993) had a group of SMEs provide information about a set of initial entry rotary-wing tasks, used the information with OSBATS to derive a set of recommendations, and then conducted group interviews with instructor pilots and researchers to determine their agreement with the recommendations; the two groups agreed with OSBATS recommendations between 70% and 98% of the time.

FORTE

Pfeiffer, Evans, and Ford (1985) developed FORTE (forecasting training effectiveness) for the Navy as a tool to estimate the training effectiveness of aviation trainers. FORTE is driven by SME data about estimated training effectiveness. SMEs estimate trials to mastery needed in an airplane by pilots with and without prior simulator training using different device features. SMEs make estimates using two different methods to check cross-method variance and rater reliability. Data may be obtained using either computer-based or hard-copy rating forms.

SMEs estimate trials to proficiency using combinations of variable conditions. Variables are treatment (experimental vs. control), student ability (fast, average, slow), task difficulty (easy, average, tough), and instructor leniency (easy, average, tough). For example, an SME is asked to estimate trials to proficiency for each set of training conditions in Table 9. Estimates are made for two groups: first for the experimental group (with prior simulator training) and second for the control group (without simulator training).

FORTE requires trials to mastery data for the 27 combinations of conditions describing the experimental group and the 27 combinations describing the control group; that is, ability (3 levels), difficulty (3 levels), and leniency (3 levels), for a total of 54 conditions. SMEs make estimates for eight conditions in each group and the remaining are estimated by computer using regression.

Pfeiffer et al. report that FORTE was validated by comparing its predictions with actual performance data:

> Validity data were obtained from the helicopter community. Estimates by flight instructors of trials-to-mastery required in the SH-3 helicopter after pretraining in

the 2F64C simulator were modified and expanded by a computer model. These modeled values were then compared with actual trials-to-mastery from a field evaluation of Device 2F64C. Reported accuracy, reliability, and concurrent validity of the model were all high and in an acceptable range (p. 5).

Interrater reliability was estimated using a number of different methods and produced values between r = .92 and r = .97. Accuracy was estimated by comparing the model's predictions with the results of a field experiment in several different test cases; in all cases, the model's predictions differed from the experiments by a few percentage points. Concurrent validity of the model was estimated at r = .80.

Table 9

FORTE Interactive Questionnaire Instrument for Estimating Trials to Mastery. (From Pfeiffer, Evans, & Ford, 1985.)

Condition	Instructor	Student	Task	Estimated trials
1	Easy	Fast	Easy	--
2	Easy	Fast	Tough	--
3	Easy	Slow	Easy	--
4	Tough	Fast	Easy	--
5	Easy	Slow	Tough	--
6	Tough	Fast	Tough	--
7	Tough	Slow	Easy	--
8	Tough	Slow	Tough	--

FORTE predicts transfer of training from simulator to aircraft. It meets the requirements for the Analytical-Transfer method-level combination in Table 5. It could conceivably be used to compare hypothetical training devices (Analytical-Comparative method-level combination in Table 5). Although FORTE was developed for use within the aviation community, it appears to have the potential for more general use.

TECIT

Goldberg, Adams, and Rayhawk developed TECIT as a CEAT tool for the Army, particularly for use in assessing TD/S (training devices and simulators) during early stages of the WSAP before empirical TEA data could be obtained. It includes both TEA and CEA models. The TEA model is described in Goldberg (1988) and the CEA model in Adams and Rayhawk (1988). The following description is based on Goldberg. The TEA model has two major components: (1) problem definition and (2) analytical forecasting and judgmental methods.

During problem definition, the analyst (with the possible aid of SMEs) defines training spectrum, context, and purpose, and gathers data and conducts baseline analyses. Objectives are to:

(1) Determine whether a TD/S is needed
(2) Aid in designing appropriate TD/S
(3) Gather baseline data on acquisition and transfer of training
(4) Provide an audit trail for applications and research
(5) Show the context and purpose(s) for which analyses are made
(6) Set the stage for designing analytic studies

Training spectrum (range of applications anticipated for the TD/S) is defined by conducting analyses and completing a written TECIT protocol. Training context (the life cycle phase of the system and related training program) and purpose are likewise defined. Information is then gathered and entered on TECIT protocols relating to weapon system, training program(s), TD/S, and predecessor and similar TD/S. Additional forms are completed to describe tasks, subtasks, and skills for the TD/S and to summarize data for analysis.

Data are then analyzed using TECIT's effectiveness function, which is defined as follows:

$$E = \left\{ \frac{S,\ ToT,\ JR}{Acq.} \right\} UR$$

E: Training effectiveness.
Acq.: Acquisition learning on the TD/S measured in terms of time to criterion.
S: Safety rating.
ToT: Transfer of training from the TD/S to an exercise on the weapon system during training.
JR: Rating of job readiness (e.g., transfer of training from the TD/S to the job.)
UR: Utilization ratio, defined as hours used divided by hours scheduled, multiplied by 100.

Factors in the numerator are combined using a weighted sum with weights based on the analyst's estimates of importance. Effectiveness increases with increases in safety, transfer, and job readiness.

TECIT provides a framework for gathering data, making estimates, and combining data to estimate transfer from training to the operational setting. Most stages of the TECIT process rely heavily on analyst or SME judgment and estimates. Because TECIT includes both TEA and CEA methods, it is a complete CEAT method. The cost-effectiveness of a particular TD/S is determined by computing a factor referred to as the operating cost ratio (OCR), which is defined as follows:

$$OCR = \frac{\text{TD/S cost/hr.}}{\text{WS cost/hr.}}$$

TD/S cost/hr. is TD/S cost per hour.
WS cost/hr. is the weapons system cost per hour.

Sticha et al. (1990) observe that comparison is straightforward when effectiveness is measured by a TER (e.g., from Roscoe, 1971), in which case cost-effectiveness is maximized by minimizing the ratio of OCR to TER. However, TECIT also uses other estimates of transfer and does not provide a complete set of rules for taking them into account.

Goldberg provided a plan for validating TECIT but it is unclear whether it was ever implemented or whether TECIT underwent other types of testing.

Like OSBATS and FORTE, TECIT predicts transfer of training from a TD/S to the operational setting. It meets the requirements for the Analytical-Transfer method-level combination in Table 5. It could conceivably be used to compare hypothetical training devices (Analytical-Comparative method-level combination in Table 5).

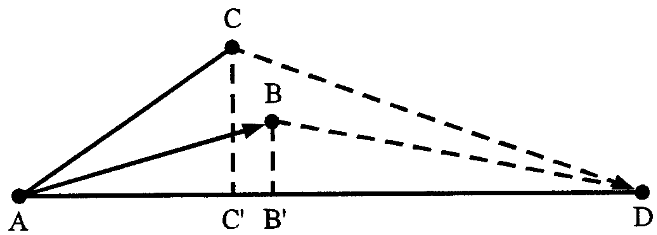
DEFT/ASTAR

DEFT is a computer-based tool for estimating device training effectiveness. It was developed for the Army and is described in a set of three reports (Rose, Wheaton, & Yates, 1985a, b; Rose, Martin, & Yates, 1985). Level of analysis can be varied with amount of information available. With very detailed information about training systems (e.g., descriptions of subtasks, displays, controls, instructional features, information about the trainee population, etc.) it is possible to perform the most detailed analysis--DEFT III (detailed subtask level) (Rose et al., 1985b). With less information, a less detailed analysis is possible--DEFT II (task level). With only general information, analysis is limited to DEFT I (global).

DEFT is based on the deficit model of training device effectiveness, which is illustrated in Figure 12 (from Rose et al., 1985a).

To apply DEFT, the analyst enters ratings on computer-based rating scales. Four analyses are required: *training problem, acquisition efficiency, transfer problem analysis,* and *transfer efficiency analysis*. After the ratings have been entered, DEFT computes several indexes and a *total effectiveness* score. These four analyses are related in pairs into acquisition and transfer components, as shown in Figure 13.

Training problem analysis estimates the magnitude of the performance deficit that trainees bring to the training device and the difficulty they will have in overcoming it.



A	= initial skills and knowledge of TRAINEE; performance on operational task prior to training on device (TD)
B	= skills and knowledge of TRAINEE at completion of TD_1 regimen; criterion performance on TD_1
C	= skills and knowledge of TRAINEE at completion of TD_2 regimen; criterion performance on TD_2
D	= skills and knowledge needed to perform operational task; criterion performance on operational equipment
B', C'	= skills and knowledge needed to perform operational task possessed by trainee after TD exposure; performance on operational equipment
AD	= time, cost associated with learning D on operational equipment
AB, AC	= time, cost associated with learning B, C on TDs
BD, CD	= time, cost associated with learning D given learning on TDs
ABD, ACD	= total time, cost associated with learning D for each TD

Figure 12. Deficit model of training device effectiveness. (From Rose, Wheaton, & Yates, 1985a.)

Acquisition efficiency analysis is conducted to describe how rapidly the training deficit will be overcome. It provides an estimate of the quality of training the device will provide to meet the training objective.

Transfer problem analysis estimates the performance deficit of trainees who have used the training device as they transition to operational equipment. Analysis estimates the size of the performance deficit and how difficult it will be for trainees to overcome it.

Transfer efficiency analysis focuses on instructional features and principles that contribute to transfer of training. It estimates the quality of training that the device provides in relation to performance on the actual equipment.

Rose, Martin, and Yates (1985) reported an analytic evaluation of DEFT indicating a high degree of interrater consistency, although no comprehensive validation was performed. With some modifications, DEFT evolved into ASTAR. A series of validations of ASTAR was conducted in the late 1980s and later a series of operational studies (Companion, 1990). Gibbons and Franchi (1990) report that the intent was to finalize ASTAR as the standard testing method in the DoD instructional system development (ISD) process, but that the outcome of the operational studies showed that ASTAR was not ready for implementation. The authors stated that ASTAR showed "promise" but that overall user acceptance was "rather low," and lacked user friendliness.

Like OSBATS, FORTE, and TECIT, ASTAR predicts transfer of training from a TD/S to the operational setting. It meets the requirements for the Analytical-Transfer method-level combination in Table 5. It could conceivably be used to compare hypothetical training devices (Analytical-Comparative method-level combination in Table 5).

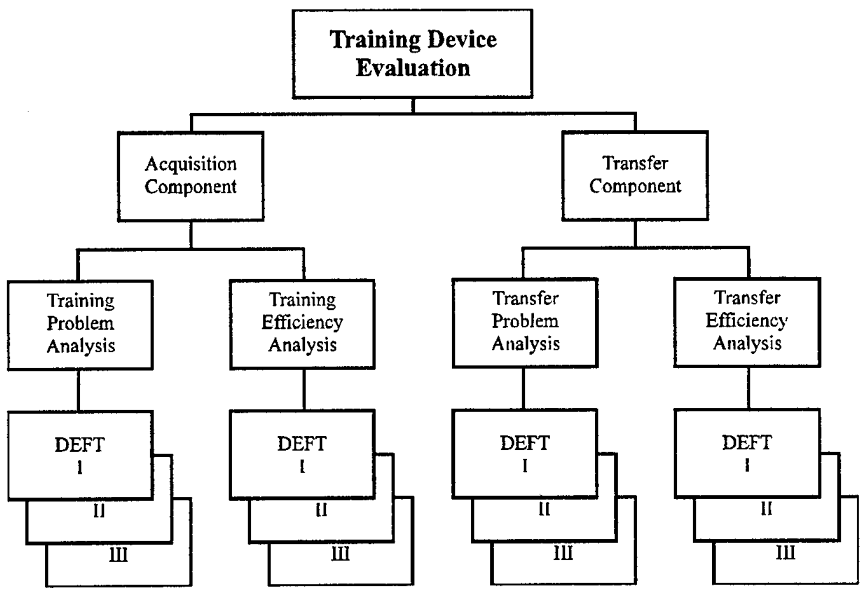


Figure 13. Types and levels of analyses in DEFT. (From Rose, Wheaton, & Yates, 1985b.)

Other Developments

The TRADOC Analysis Center, White Sands Missile Range (TRAC-WSMR), recently developed an analytical method to evaluate the cost and training effectiveness of various mixes of field training and training using TADSS (training aids, devices, simulators, and simulations). The "training mix model," as it is known, is a mathematical programming model that incorporates the expected cost of acquiring and using training systems with their expected effectiveness in terms of ability to train required tasks (Djang, Butler, Laferriere, & Hughes, 1993). Although this method bears some similarity to OSBATS, it was developed independently at TRAC-WSMR and has no familial linkage to OSBATS or other prominent analytical methods. The method was described in the 1993 report along with several examples of its applications. No validation data were provided although the authors considered the method promising and indicated that it would undergo further development.

The Simulator Systems Research Unit of the ARI, co-located with the STRICOM (U.S. Army Simulation, Training, and Instrumentation Command) in Orlando, recently conducted a study for STRICOM to develop an analytical method that could be used to predict the cost-effectiveness of a training system during the conceptual stages of system design (Witmer, 1991). Witmer analyzed six existing methods (TRAINVICE, DEFT, TECIT, TEEM, CBP, and OSBATS) in terms of their scope and ability to provide cost estimates. Ultimately, he conceptualized and described a new method, VALTRAIN (value of training), which is intended to overcome perceived limitations of the methods reviewed. VALTRAIN presently exists as an interesting concept presented in a report; it has not yet been applied. Its outlook is doubtful, as STRICOM is not currently investing significant resources in research relating to analytical methods.

What the SMEs Said about Analytical Methods

Interviews with SMEs at ARI and HUMRRO indicated that none of the five key methods described in this report is currently in general use. Further, there appeared to be consensus that (a) development of analytical methods has languished in recent years due to lack of resources, (b) methods are often perceived by users to be difficult to apply and to lack "user friendliness," (c) methods lack validation by comparison of their results with empirical methods, and (d) proponents often find it difficult to convince military decision-makers that analytical methods produce valid results. While acknowledging these serious problems, most of the SMEs who had experience conducting or overseeing research with analytical methods believed that further research and development with them was warranted.

The most recent review of analytical methods was published in 1994 by Muckler and Finley, who compared and summarized the attributes of 36 different training system estimation models (including all of those described in the present report). The authors acknowledged problems with the methods but took a somewhat charitable view of the field:

> To a great extent, the last 20 years have been a period of trying ideas, some of which have been very complex. In one sense, the "state of the art" is a large learning experience from which many major future advances may be possible. One point of view is that the past two decades were necessary to structure and to begin to understand the problem (p. 4).

Elsewhere, the authors commented on the lack of adequate documentation concerning the methods, noting that few of the methods had gone beyond the research stage, few review articles had appeared, and that in about 30 percent of cases, no formal reports or documentation were obtainable. The authors recommended that some organization or professional society should institute an archival store of documentation so that past mistakes would not be repeated and past successes might be built upon.

Cost Analysis

Life Cycle Costs

Cost analysis (CA) is conducted to assess the resource implications of the alternatives being considered in CEAT. One of the basic notions of CA is life cycle costs. *DoD Instruction 5000.2* (DoD, 1991) defines life cycle cost as follows:

> Life-cycle cost reflects the cumulative costs of developing, procuring, operating, and supporting a system. They are often estimated separately by budget account (i.e., research, development, test, and evaluation..., procurement, and operations and maintenance). It is imperative to identify life-cycle costs, non-monetary as well as monetary, associated with each alternative being considered (p. 4-E-4).

Costs accrue over the life of a system. TRADOC divides this life into five distinct but sometimes overlapping phases (Department of the Army, 1985):

- Conceptual (exploratory development): Solicitation, evaluation, and exploration of alternative concepts.
- Demonstration and Validation (advanced development): Prototypes are produced to support demonstration.
- Full Scale Development (engineering): Prototypes are produced to support operational test and evaluation
- Production and Deployment
- Operation and Support

Cost Element Structure

Knapp and Orlansky (1983) developed a comprehensive cost element structure which has become widely accepted and used as a framework for cost estimation over the life cycle of training programs, courses, and devices (P/C/D) of varying complexity. It was derived from several authoritative and widely used cost guides and was reviewed by nearly 50 government representatives concerned with costing before being published. The structure is organized in the form of an outline. The top two levels of the outline are shown in Figure 14.

In principle, the cost element structure provides an inventory of what to consider when making a cost estimate. One then determines what elements to include and their costs. In practice, it is not this simple for a number of reasons. First, it may be difficult or impossible to obtain some of the cost data directly. Particularly with new systems, these data may simply be unavailable. When this is the case, costs will have to be estimated. There are several methods for doing this (as discussed in greater detail in the *CEAT Concepts* section of this report).

A. RESEARCH AND DEVELOPMENT

1. Design
2. Component Development
3. Producibility Engineering and Planning
4. Tooling
5. Prototype Manufacturing
6. Data
7. Training P/C/D Test and Evaluation
8. System/Project Management
9. Facilities
10. Other

B. INITIAL INVESTMENT

1. Production
2. Engineering Changes
3. Purchased P/C/D - Peculiar Equipment
4. Common Equipment
5. Data
6. Training P/C/D Test and Evaluation
7. System/Project Management
8. Rents
9. Operational/Site Activation
10. Initial Training
11. Transportation
12. Other

C. OPERATING AND SUPPORT

1. Direct Costs
2. Indirect Costs

Figure 14. Top two levels of Knapp and Orlansky's (1983) cost element structure for defense training.

Economic Factors

Cost estimation should take into account several economic factors. The following is distilled from a description of key factors in Adams and Rayhawk (1987):

- *Opportunity Cost vs. Accounting Cost*: Accounting cost is the cost "on the books." Opportunity cost is the hypothetical value of a resource in its "best alternative use."
- *Sunk Costs*: Costs that have already been incurred and that cannot be recouped. An example is the cost of R&D spent on various forms of technology.
- *Fixed and Variable Costs*: Fixed costs are not affected by how much training occurs; an example would be the cost of classroom space. Variable costs vary with the amount of training; an example would be the cost of instructors, whose number would vary with the student load.
- *Time Value of Money*: The value of money changes with time because money has earning power. This is considered when comparing alternatives whose expenses are incurred at different rates over various periods of time by estimating both costs in terms of "present value" dollars.

- *Discount Rates*: Costs that can be deferred into the future can be discounted because a smaller amount of money could be invested today and earn interest to make the future payment.
- *Constant vs. Current Dollars*: The purchasing power ("current value") of the dollar varies with the general price level and inflation rate. Constant dollars reflect the purchasing power of the dollar in a selected base year.
- *Residual Value of Assets*: The value, if any, left after a system has completed its life; for example, the value of computers, part of whose cost may be recouped.
- *Indirect Benefits*: Benefits which may occur beyond the intended scope of training. For example, the value of training to individuals in preparing them for a civilian occupation. These are not usually considered in CEAT.

Sources of Cost Analysis Data

Knapp and Orlansky's cost element structure is an inventory of what elements to include in cost analysis but does not provide cost data for assigning costs to the elements during a CEA. The process is analogous to valuing MOE during a TEA. Some of the data sources commonly used:

- Compute current cost estimates
- Estimate cost based on historical data
- Estimate cost based on similar systems
- Obtain SME estimates
- Develop and evaluate analytical methods

As with MOE, some of these methods (the first four) rely on observation and experience and formally are empirical methods. Undoubtedly, they vary in credibility, though in ways best left to cost experts to judge.

There is also a class of analytical methods for cost estimation. These will not be covered in this report. Good reviews of cost methods are presented in Adams and Rayhawk (1987, 1988). The Litton cost model, an attempt to integrate several cost models existing in the early 1980s, is described in Matlick, Berger, Knerr, and Chiorini (1980) and Matlick, Berger, and Rosen (1980).

Sensitivity Analysis

The objective of a sensitivity analysis is to determine how sensitive study conclusions are to changes in the important variables driving the analysis; this facilitates more reliable decisions and error estimates (Swope, 1976).

DoD Instruction 5000.2M (DoD, 1991) says the following about cost sensitivity analysis:

> Cost sensitivity is the degree to which changes in certain parameters cause changes in the costs of a system. Each potential change should be tested independently. Operating parameters that affect costs (such as activity rates and performance

characteristics) should be examined for sensitivity to change. The results of each sensitivity analysis must be documented (pp. 8-9).

Obviously, the context for these analyses is COEA rather than CEAT but the guidance and basic principle both apply.

The *CEAT Methods* section described sensitivity analysis within the context of CBA, pointing out that the assumptions underlying a CBA lead to uncertainty in the outcome of analyses; this is equally true for CEAT. If the CEAT is locked into a single set of assumptions with the intent of obtaining a definitive result, its outcome may be untrustworthy. Hence, sensitivity analysis varies the assumptions systematically to provide the results of analyses under different sets of assumptions. As pointed out in the discussion of Figure 1, sensitivity analysis modifies the assumptions and recycles the CEAT process iteratively to "tune" analyses to reveal the best training alternative.

This report does not discuss sensitivity analysis further as these methods are fairly well documented elsewhere. Sassone and Schaffer (1978) provide good explanations of analysis within the context of CBA. Several TRADOC publications cover analysis within the context of COEA (e.g., *TRADOC Pamphlet 11-8: Studies and Analysis Handbook* and *Methodology for Abbreviated Analyses*).

Choosing among CEAT Methods: A Dilemma

The cost analysis part of CEAT is fairly well defined. It is not simple, but it is arguably not as difficult or complex as the TEA part. Performing a TEA poses at least two problems: (1) deciding what type of TEA to perform and (2) actually performing the TEA. Consider the first problem. Table 5, relating TEA evaluation methods and levels, suggests 16 different classes of TEA. Hence, there are several times 16 ways to perform a TEA or CEAT.[12] How does the analyst decide what type of CEAT to perform?

In many cases, the choice of method is probably taken for granted; the analyst may decide that the method to use in the present case is the same one that was used in the last similar case. Precedent and familiarity govern the options considered. While this may lead to reasonable choices, it will not always lead to the best choice. All reasonable options should be considered because the most appropriate method to use in a particular situation should be influenced by such factors as the purpose and objectives of the analysis, methodological validity and reliability, cost, and analysis requirements and constraints.

Time

Figure 9 and the related discussion in the *CEAT Methods* section illustrated that time affects the availability of the data necessary to conduct a CEAT. At the outset, no training system exists and no empirical training data are available. As the system is developed, data become

[12] Although all of the classes of TEA shown in Table 5 are hypothetically possible, some are much more likely to be used in practice than others.

increasingly available, but the ability to change the system declines rapidly. By the time it is possible to run experiments, it may be difficult or impossible to change the system and, the more costly the system, the less freedom one has to abandon a particular training system in favor of another possibility. TEA methods that do not rely on experiments--SME-based, comparison-based, and analytical--have the potential to generate data earlier in the WSAP than do experiments. Hence, these methods may be used because there is no other source of data at a particular point in time. In any of these cases, enough must be known about the system and the tasks its users perform to meet the information requirements of the particular method.

Validity and Reliability

The four general classes of methods discussed earlier in this section (experiment, comparison, SME estimates, analytical method) vary in terms of validity and reliability and in turn in their credibility; credibility reflects the general faith that science attaches to the validity and reliability of the methods. The reliability of the scores generated by the four kinds of methods limits the validity of inferences that can be drawn from those scores. Reliability and validity in turn influence the credibility of results.[13] In this race, experiment has traditionally won, as it is the method of choice in the research laboratory and is generally perceived as having the greatest face validity. The credibility of data based on similar systems or SME estimates depends upon the situation but, in general, is somewhat lower. The validity and reliability of analytical methods is for practical purposes unknown because so little validation has been performed, although it is clear that analytical methods lack credibility for this (as well as other) reasons within the military community. Where DoD takes a stand on this matter, it tends to be on the side of experiment, as most DoD and Service written guidance on TEA (see the *CEAT Written Guidance* section of this report) either directs or implies that field trials should be performed.

Is there anything wrong with this picture? If in conducting a TEA one can approach the level of control obtainable in the scientific laboratory, then the experimental bias is justifiable. However, if one cannot, then blind faith in experiment is misguided. TEAs conducted in the operational setting commonly lack the controls necessary for valid causal inferences about training effects. Boldovici (1987) noted that these kinds of studies often are compromised by too few subjects (hence inadequate statistical power to demonstrate differences between scores of compared groups), uncontrolled pre-experimental differences between compared groups,

---

[13] Boldovici (1995) contends that, because analysts do not typically report the reliability of scores generated by the various methods, there is no objective basis for judging the relative validity of inferences drawn from the results of different methods. In his view, in the absence of that objective basis, the results of experimental methods seem to be preferred to the results of the other methods, perhaps because of "analysts' and evaluators' failure to inform [military] leadership about the difference between form and substance in experiments." Boldovici pointed out a few cases where reliability was reported. Powers, McCluskey, Haggard, Boycan, and Steinheiser (1974) reported the split-half reliability of the results of their live-fire tank-gunnery scores to be no better than random guessing. Boldovici (1995) contrasted that estimate of the reliability of field trial scores to Burnside's (1990) and the U.S. Army Armor School's (1989) reports of SME-rating reliabilities for trainability of tasks with SIMNET. Noting that the SMEs' ratings ranged from 72% to 98% agreement depending upon methods used, Boldovici concluded, "The sparse evidence bearing on the potential validity of inferences from SME methods and from field trials with Army training devices suggests that the SME methods win hands down."

confounded treatments (e.g., kinds of training confounded with amounts), and other factors.[14] In a review of several costly field studies, Boldovici and Bessemer (1994) noted that all were so flawed as to preclude valid inferences about the effects of training with the device of interest (SIMNET) on unit proficiency in the field. The authors concluded that SME-based methods provided useful diagnostic information at much lower cost than the flawed field trials. The authors made the point that SME-based estimates provided useful information at much lower cost than the types of flawed field experiments that were common.[15] Hence, the picture is more complicated that it seemed at first.

One can argue that a flawed experiment is not really an experiment, but rather a futile exercise. However, if the available resources preclude the conduct of a proper experiment, what would one expect a real-world analyst to do? The answer is left to the reader.

The previous paragraphs contrasted experiment and SME-based methods. Further elaboration is possible regarding the relative merits of comparison-based methods in this equation, as well as all the different classes of methods. At this point, it is safe to exclude analytical methods from consideration as none is in widespread use.

Cost

Cost has two aspects: (1) cost of the system being analyzed and (2) cost of conducting the analysis. DoD directives set cost thresholds which govern when analyses are to be performed (see the *CEAT Written Guidance* section of this report). The guidance is somewhat ambiguous, but a liberal reading of the directives suggests that CEAT is not normally required unless a training system represents an RDT&E investment much greater than roughly $100M. Very few training systems cost this much, so the training developer normally has little incentive--other than a genuine interest in CEAT--to perform analyses.

The cost to perform a CEAT varies with the method used and the situation. As noted previously, cost is generally highest for experiment and lowest for SME estimates, with estimates based on similar systems in the middle. On this point, Boldovici (1995) made the following observations:

---

[14] Poorly controlled field trials often yield no statistically significant difference between the scores of compared groups, a result that may ensue from inadequate statistical power. The null result may then be misinterpreted to mean that the two groups are equivalent. However, the null result does not establish equivalence; it only renders one unable to say that the scores of the compared groups differed. Boldovici (1995) noted that examining the equivalence of compared groups' scores requires using power analysis and confidence intervals and added he had seen neither reported in any field test of an Army training device. Additional explications of this view are in Cohen (1990, 1994) and Festinger and Katz (1953). The author is indebted to John Boldovici for bringing this material to his attention.

[15] Boldovici (1995) argues that (1) the common objection to SME-based data because of their subjective basis should also apply to data obtained in field trials where scores represent the judgment of SMEs acting as observer-controllers; (2) to date, the reported reliabilities of scores are greater for SME-based methods than for field trials, and hence the potential validity of inferences is greater for SME-based scores; (3) the validity of SME-based methods cannot be established via comparison of results with a one-shot field trial; and (4) the generality of field-trial results can only be known in light of replication, which is not feasible for multi-million dollar device tests.

> Parsimony demands that the prices of SME-based methods vs. the prices of quasi-experimental field trials figure in our choice of methods. Burnside's SME-based SIMNET analysis cost approximately $50,000 and Drucker and Campshure's approximately $100,000. The GAO reported estimates of $15M to $19M for operational testing of the Close Combat Tactical Trainer. Assuming [erroneously, according to Boldovici] that the results of SME-based methods and field trials are equally unreliable, we should rather spend $50,000 than $19M (p. 16).

Clearly, the range of costs is enormous, and the analyst cannot ignore cost in picking a method. In fact, the analyst is required to make a cost-effectiveness tradeoff in deciding what type of CEAT to perform. The question being asked is, "How cost-effective is it to conduct each possible type of cost-effectiveness analysis?"

Analysis Requirements and Constraints

In the perfect world, the cost of performing CEAT is not a concern, there are no time limitations, the military organizations that will participate in testing willingly offer complete cooperation during data collection, and the designated analysts have professional training and are highly skilled and experienced.

The real world is of course not like this. Cost and time both limit what types of analyses are feasible. The operating schedules of military organizations limit their ability to participate in testing, the levels of training of participating military personnel will vary (and may change during testing), and willingness to cooperate during testing cannot be guaranteed. Finally, the analysts who will perform CEAT vary in terms of training, skills, and experience. All of these factors have an impact on the quality of the CEAT performed and on the possible outcome.

Making the Choice

The preceding paragraphs describe some of the important factors to consider before performing a CEAT. At present, it does not appear that these and other appropriate factors are formally considered before performing a CEAT. Even if the analyst wanted to take these (or other) factors into account, this study did not reveal any systematic method for making the choice.

CEAT WRITTEN GUIDANCE

This section reviews written guidance produced within the DoD, the Services, and by contractors since 1980. The documents are believed to represent most of what has been written on the subject of when and how to perform CEAT during the time frame covered, but the literature review could easily have missed some important items. Several TRADOC documents which appeared to relate to CEAT were impossible to obtain as they were out of print or undergoing revision and were unavailable from the usual sources. This section discusses DoD, Army, and other guidance. It does not cover guidance relating to the analytical TEA methods described in the *CEAT Concepts* section of this report (OSBATS, ASTAR, FORTE, TECIT) as it does not appear that any of the methods is used today by the military to evaluate training systems.[16]

DoD Guidance

DoD Instruction 5000.2: Defense Acquisition Policies and Procedures (DoD, 1991), Part 4, Section E ("Cost and Operational Effectiveness Analysis"), provides policies and procedures for conducting COEA. COEA are required for all acquisition category I programs; these are programs designated by the Under Secretary of Defense as category I or that involve an expenditure for RDT&E of $300M or more or a total procurement cost of more than $1.8B in 1990 constant dollars. CEAT is not specifically mentioned. If one infers that CEAT is the form of COEA required when a training system is being developed, the instruction could be interpreted as requiring CEAT, although few training systems cost anywhere near this much. Even if many did, it would be reasonable to conclude that the instruction does *not* require CEAT. The guidance is brief and general. The instruction essentially lists and describes some of the elements in COEA (e.g., mission need analysis, threat vs. U.S. capabilities, the use of MOE, some basic cost elements and considerations in cost-effectiveness comparisons, role of OSD, and milestone decision reviews). The guidance helps clarify DoD COEA requirements, explains some basic concepts, but does not tell how actually to perform the work.

DoD Instruction 5000.2M: Defense Acquisition Management Documentation and Reports (DoD, 1991), Part 8 ("Cost and Operational Effectiveness Analysis"), covers much of the same ground as 5000.2, in slightly more depth. As with 5000.2, CEAT is not mentioned, it is ambiguous whether the instruction requires CEAT for training systems, and there is a lack of how-to guidance. Notably, the instruction stresses the importance of considering alternatives in conducting COEA, and mentions modeling as a way to predict how a system would work. Both of these are important considerations in conducting CEAT and the instruction puts the DoD on the record regarding the use of modeling, when appropriate, and using analytical techniques that compare alternatives (as opposed to less stringent testing requirements).

DoD Directive 1430.13: Training Simulators and Devices (DoD, 1986) establishes DoD policy for acquisition of training simulators and devices and may reduce slightly the ambiguity in 5000.2 and 5000.2M regarding CEAT requirements. Among other things, the directive requires that when a Service considers acquiring a training device it shall conduct an analysis to "evaluate the benefits and tradeoffs of potential alternative training solutions" (p. 3). Explicitly, the directive covers embedded training, training devices and simulators, and training systems. The directive states that economic analyses should be conducted, where applicable. Of training effectiveness evaluation, the directive offers one sentence: "Analysis of training capability and potential should focus on data based on actual experience" (p. 4). It is not clear whether this should be taken to mean experiment only or whether it also includes other empirical methods. The statement seems to exclude analytical methods, including the "models" mentioned in 5000.2M. The directive does not mention CEAT, is quite general, is somewhat ambiguous on whether it includes certain training technologies (e.g., distance learning technologies, training

[16] Readers interested in guidance on analytical methods should refer to Muckler & Finley (1994a,b) as well as other documents cited in Chapter 3.

delivery media), and imposes a reporting requirement only if cost rises above a threshold[17] or the Secretary of Defense expresses special interest in the device.

MIL-STD-1379D: Military Training Programs (DoD, 1990) applies in all military training program acquisitions and major modification programs. Nothing in this document requires that CEAT be performed. Task sections 100, 200, and 300 provide boilerplate-type task statements that can optionally be included when contractors are hired to develop training systems. One of these tasks, 206, "Training System Alternatives Identification," calls for the contractor to "evaluate each alternate, in terms of cost, relative to its capability to meet training constraints and requirements [and to] identify the best suited alternate" (p. 75). If the requirement is included, the method used to satisfy it is open to interpretation.

The systems approach to training (SAT), also known as the method of instructional system development (ISD), was adopted as the standard method for developing instruction in the DoD many years ago. *MIL-STD-1379D* sketches its five-step method (Analyze-Design-Develop-Implement-Evaluate) and the Services have developed their own implementation documentation. The Army's SAT guidance, which is typical, is *TRADOC Regulation 350-7: Systems Approach to Training* (Department of the Army, 1988b). Notably, SAT/ISD does not explicitly deal with CEAT concepts such as training alternatives, relative costs, cost-effectiveness, etc. Evaluation is an important part of SAT as a way to continuously validate instruction. Yet the method itself seems to view all training as classroom-based, using traditional methods, and relatively static. This method, developed in the mid-1970s, needs to be updated to include cost-effectiveness considerations.

Army Guidance

TRADOC Regulation 350-32: The TRADOC Training Effectiveness Analysis (TEA) System (draft) (Department of the Army, 1993) establishes TRADOC policies, procedures, and responsibilities for TEA studies. An update of the 1990 version of the regulation, it uses simplified terminology to combine such terms as CTEA, DTEA, PFTDS, TEA, etc. under the single umbrella term TEA. (TEA fits this report's definition of CEAT.) The regulation states, "TEA studies provide cost and effectiveness information," employ "qualitative and quantitative analytical techniques to derive information" (p. 10), and are conducted for purposes related to (a) system acquisition, (b) resolving training problems, or (c) improving study methods. It shows clearly and graphically when TEA are required in terms of the phases of the WSAP.

TRADOC Training Effectiveness Analysis Handbook (first draft) (Department of the Army, 1980) "is a guidance document for planning, conducting, and writing...TEA" (p. vi). In 11 chapters and 12 appendices it describes the types of TEA, how they fit into the WSAP, when they are required, roles and responsibilities for conduct, and methods for performing studies. Three chapters deal with method: 8 (guidance for design and analysis), 9 (data collection instruments), and 10 (cost analysis). These chapters identify and briefly cover technical matters such as literature search, sample selection, statistical analysis, study design, data collection methods,

[17] The actual thresholds are given in Part II of *DoD 7110.1-M: Budget Guidance Manual*, 1985. This directive is currently out of print.

acquisition of cost data, etc. The impression is given that the material was intended for use as a cookbook for performing TEA by an audience that lacks sophistication in the area.[18] The approach described is exclusively experimental and comparative (C-E combination in Table 5). Despite any philosophical reservations one may have about the handbook's approach, it provides much specific and usable guidance for conducting CEAT. Ironically, it was never published and continues to be available only in draft form; as such, it does not provide formal guidance.

TRADOC Pamphlet 11-8: Army Programs: Studies and Analysis Handbook (Department of the Army, 1985) provides general guidance for planning and conducting several different types of studies required by *Army Regulation 5-5* (Department of the Army, 1981). The focus is on COEA, and CEAT is not explicitly covered, but much applies to CEAT (e.g., sensitivity analysis, experimental design, cost analysis, study plan for conducting COEA, method for making cost-effectiveness tradeoff.

Methodology for Abbreviated Analyses (Department of the Army, 1986) describes how to conduct abbreviated COEA, which the Army permits under certain conditions for non-major system acquisitions. The focus is on COEA, and CEAT is not explicitly covered, but much applies to CEAT. The method starts with a list of alternatives to be compared. Data are then obtained or estimated during these steps:

- performance analysis
- cost analysis
- comparison of alternatives
- sensitivity analysis

The performance data used are the best that can reasonably be obtained, and may range from actual raw data to verbal estimates. (The method could be applied to training by estimating training effectiveness rather than system performance.) The method is comparative and is probably most likely to use analyst or SME estimates (SME-comparative method-level combination in Table 5). The procedure guides the analyst through a data gathering, analysis, and decision-making process that leads to a recommendation. This is a useful analytical framework--not as good as a controlled experiment, but better than unstructured estimation.

In addition to the foregoing, according to a recent briefing (TRADOC, 1993a), the Army has published several other guidance documents:[19]

- *AR 71-9: Materiel Objectives and Requirements* (under revision)

---

[18] The notion that it is possible to accomplish this might be greeted skeptically in graduate schools. People who perform CEATs deal with many complex methodological issues--experimental design, statistical analysis, sample selection, developing data collection instruments, etc. Such topics are typically covered in universities at the graduate level if the student intends a career in scientific research. Graduate students usually serve apprenticeships for faculty to apply research skills and are expected to demonstrate them to receive a graduate degree. Hence, the idea that one can shortcut this process by providing people with a written guide would bother many experts who are not in the DoD.

[19] The author was unable to obtain copies of these documents for review.

- *AR 70-1: Systems Acquisition Policy and Procedures* (under revision)
- *AR 350-38: Training Device Policy and Plans* (1992)
- *DA Memorandum, Subject: COEA Policies, Procedures and Responsibilities,* July, 1991 (interim COEA policy)
- *TR 351-9: Systems Training Development* (1989)
- *Non-System Training Device Study Process*

Other Guidance

In 1993, Derrick and Davis of the Air Force performed a CEA of the C-130 air crew training system for the Armstrong Laboratory. Their research report was not intended to be used as a how-to guide but provides such a detailed explanation of its method that it could probably be used for that purpose. In considerable depth, the report describes key CEAT concepts (e.g., CEA framework, cost analysis, cost elements, performance measurement, sensitivity analysis) and then applies them to the analysis of the C-130 air crew system. The authors developed cost analysis spreadsheets, which are available upon request. The analyses are well enough explained to be useful examples for others.

The University of Central Florida was contracted to develop CEAT guidance and produced a two-volume report: Vol. I (Hall, Kincaid, Muller, & Kiernan-Kostic, 1994); Vol. II (Hall, Kincaid, Braby, Kiernan-Kostic, Muller, & Walker, 1994). Volume I (Acquisition of Data) describes how CEAT fit into the DoD procurement process to comply with milestone, ISD, and *MIL-STD-1379D* requirements. It suggests how to use existing DoD mechanisms to meet mandated requirements. Volume II (Procedural Guidance) sketches a general CEAT method based on *TRADOC Pamphlet 71-10* and briefly elaborates each step. Volume I draws together much information that is available elsewhere but breaks no new ground. Volume II elaborates an existing CEAT framework but at such a general level and with such vague guidance that it is not really clear what, exactly, an analyst needs to do to perform a CEAT.

Calspan Corporation developed a three-volume guide for conducting training effectiveness evaluations (TEE) for air defense training (Fishburne & Rolnick, 1985; Larsen, Rolnick, & Fishburne, 1985; Rolnick, Fishburne, & Nawrocki, 1985). The guide presents a detailed, step-by-step method to design a TEE, develop data collection instruments, collect and analyze data, and diagnose and correct training program deficiencies. In overview, the method consists of six steps: plan TEE, conduct product evaluation, plan training process, conduct training process evaluation, assess trainee performance, and document TEE. The accompanying guides describe the process in detail and provide numerous examples. The method is similar to IQI in that it uses a structured process to evaluate existing training in terms of certain attributes that should be present. It is driven by a combination of empirical data (e.g., written and performance test scores) and analyst/SME data (e.g., evaluation checklists). It appears that the authors intended the method to evaluate training against a set of standards (i.e., non-comparatively). The method could conceivably be used to compare two different courses, but to do so both courses would have to exist in mature form.

Two other methods, both described in the *CEAT Methods* section of this report, should be also be mentioned under this heading as the authors have provided written guidance explaining how to apply them: IQI, and Frederickson's SME-based method. Refer to *CEAT Methods* for additional information.

Finally, it should be noted that for the period 1970-1985, the Chief of Naval Education and Training (CNET) maintained the Training Analysis and Evaluation Group (TAEG), located at the Naval Training Systems Center (NTSC). TAEG published several reports dealing with training evaluation, cost analysis, and related topics. TAEG later merged with NTSC and no successor organization has carried on its work or continued to publish in the general area of CEAT.[20] Several TAEG reports remain useful today and are often cited in the training literature:

- *Staff Study on Cost and Training Effectiveness of Proposed Training Systems* (Braby, Morris, Micheli, & Okraski, 1972)
- *A Primer on Economic Analysis for Naval Training Systems* (Swope, 1976).
- *A Technique for Choosing Cost-Effective Instructional Delivery Systems* (Braby, Henry, Parrish, & Swope, 1975)
- *Training Effectiveness Assessment: Volume II, Problems, Concepts, and Evaluation Alternatives* (Hall, Rankin, & Aagard, 1976)
- *Modeling Field Evaluations of Aviation Trainers* (Pfeiffer, Evans, & Ford, 1985)

CEAT IN THE SERVICES

This section provides an overview of how the Services deal with CEAT. The information presented is based on the literature review and discussions with several training SMEs in the Services.[21] Whether or not a Service has a CEAT "program" depends upon one's definition of the word. For purposes of this discussion, a program is defined by three indicators: (a) the Service promulgates in writing the requirement to perform CEAT, (b) a Service organization exists that is formally tasked with performing CEAT, and (c) the Service publishes CEAT reports.

---

[20] The organization still widely referred to as NTSC has undergone several reorganizations and name changes. Originally known as the Special Devices Center (located at Great Neck, Long Island), it became, in order, the Naval Training Device Center (NTDC), Naval Training Equipment Center (NTEC), NTSC, and most recently, the Naval Air Warfare Center, Training Systems Division (NAWCTSD).

[21] The author makes no pretense that the information presented here is comprehensive. It would have been safer to omit this section than to present information with a caveat, but the subject is important and even the limited information presented is not readily available elsewhere. To capture all the details of CEAT in each Service would require a large scale survey that was beyond the scope of the present study. This section presents an analysis based on admittedly limited data. No empirical data were gathered during the study to indicate that any Service has consistently performed CEAT in an outstanding manner or failed to perform CEAT when appropriate. The study found apparent differences among the Services in how they deal with CEAT. The Army has a much more structured and centralized approach than the Navy or Air Force. Proponents of less structured and decentralized CEAT might argue that the Army's approach is rigid and bureaucratic. It is left to the reader to decide whether the Army should, or should not, be used a model for the other Services and whether the apparent differences among the Services are of any practical significance.

Army

The Army meets all the definitional requirements for having a CEAT program. It has published regulations and guides for conducting CEAT (see the *CEAT Written Guidance* section of this report). TRADOC Analysis Centers (TRAC) are formally tasked with performing analyses. And the TRAC regularly publishes CEAT reports. CEAT in the Army appears to be organized and under centralized control.

According to a recent briefing (TRADOC, 1993a) TRADOC' s TEA program conducts studies to assess training effectiveness and costs of TRADOC training strategies, programs, and products throughout the five phases of the ISD process (analyze, design, develop, deliver, evaluate) in accordance with the DoD, Army, and TRADOC requirement and guidance documents described in *CEAT Written Guidance*. Studies focus on cost-effectiveness relating to acquisitions, resolving problems with fielded training programs and products, and cost-effectiveness of training innovations.

The *FY94 TRADOC Study Program Study Descriptions* (TRADOC, 1993b) lists approximately 200 studies underway or about to be conducted as of the start of FY94. The majority of studies dealt with non-training related issues but approximately one-fourth are related to training. The training studies concerned training on equipment (e.g., Bradley), training with devices (e.g., aviation simulators), and generic training (e.g., improving an NCO course).

Navy

The Navy has not published regulations or guides for conducting CEAT, does not have an organization formally tasked with performing such analyses, and does not regularly publish CEAT reports. By the definition being used, the Navy lacks a CEAT program.

OPNAV Instruction 5311.7: Determining Manpower, Personnel and Training (MPT) Requirements for Navy Acquisitions (Department of the Navy, 1985) established a requirement for the Hardware/Manpower Integration (HARDMAN) program to be implemented Navy wide, and procedural guidance was subsequently provided in *The Navy Program Manager's Guide to Early MPT Planning* (Department of the Navy, 1987). While not a CEAT program, HARDMAN provides a comparison-based method for predicting the impact of a program on manpower, personnel, and training (MPT). The method considers costs and program alternatives so arguably it provides some information of the type required by CEAT.

It appears that CEAT is not performed routinely in the Navy, although it may occur when a perception of need arises. Navy systems are developed by Navy system commands (SYSCOMs), whose orientation is primarily toward hardware and software development rather than training. A SYSCOM may determine that CEAT or other analyses are needed, and often it will obtain analytical support from NAWCTSD. NAWCTSD itself may perceive the need for analysis and approach the SYSCOM to obtain funding to conduct the analysis. In some cases, CNET may perceive the need for analysis and direct that it be performed. In all of these cases, the decision to conduct analysis is driven on an ad hoc basis by a perception of need rather than by Service CEAT policy and program.

One noteworthy example of CNET involvement in CEAT occurred in 1989-91 as the Navy considered adopting video teletraining (VTT). A program begun in 1988 as a demonstration project appeared to be operating successfully but CNET raised questions about its cost-effectiveness. At CNET urging, the NTSC conducted analyses of VTT costs and potential applications (Sheppard, Hassen, Hodak, Swope, & Denton, 1991; Sheppard, & Hodak, 1991). The Center for Naval Analyses (CNA) was engaged to conduct a study to assess system utilization, training effectiveness, downtime, and savings to the Navy (Rupinski & Stoloff, 1990). Among other things, the study compared the performance of students at VTT sites where the instructor was physically present (representing a control condition) to performance at remote sites (experimental condition) and the costs of VTT vs. traditional live instruction. This study, though hampered by many of the problems commonly encountered when performing evaluation research in the operational setting, met many of the requirements of a CEAT. Its authors contend that it demonstrated the cost-effectiveness of VTT. It appears that all of these analyses were motivated not by DoD or Service requirements but because CNET judged it necessary to justify further expenditures on VTT.

Air Force

The Air Force has not published regulations or guides for conducting CEAT, does not have an organization formally tasked with performing such analyses, and does not regularly publish CEAT reports. By the definition being used, the Air Force lacks a CEAT program.

It appears that CEAT is not performed routinely in the Air Force, although it may occur when a perception of need arises. The Air Force is the DoD's largest user of flight simulators and has conducted many studies to justify their cost-effectiveness. Perhaps a fraction of these studies have been published in a form that they are available for general access. Another example of Air Force work is analyses performed in connection with the acquisition of contractor-based training programs. In recent years, the Air Force has contracted much of its training; currently it contracts training of air crews for several aircraft as well as undergraduate pilot training. Cost is agreed to under the terms of the contract. Training effectiveness evaluation is performed jointly by the contractor and the Air Force, typically evolving through formative evaluation early on, summative evaluation after a contract is signed, and operational evaluation once a training system is in place. Hence, the training system acquisition process deals with both cost and training effectiveness issues, though not in the formally structured manner that occurs in the Army.

A noteworthy example of Air Force CEAT is a project conducted by Systems Research and Applications Corporation (under contract to the Air Force Armstrong Laboratory) to perform a landmark training cost-effectiveness study which compared traditional (Air Force-conducted) training with contractor-conducted training for the C-130 air crew training system (Derrick & Davis, 1993). The study demonstrated the cost-effectiveness of the contractor-based training approach. In addition, the contractor developed and provided a computer-based automated cost model which generalizes to other types of training systems. As noted in the *CEAT Written*

Guidance section, the methodology used was exceptionally well presented and could be used as an example by others. The Derrick and Davis study probably does not typify the way CEAT is conducted in the Air Force because the system it focused on had significant cost implications for the Air Force. Nonetheless, the study illustrates that effective CEAT can be conducted without a centrally-controlled CEAT program in place in the Service.

CONCLUSIONS

Objectives of the study were to:

- Determine the current state of knowledge and research on conducting CEAT
- Identify documented CEAT methods
- Develop a CEAT general conceptual model
- Assess the current status of CEAT in the Services
- Determine potential areas where R&D would be useful

Previous sections of this report present the information and analyses intended to meet the first four of the objectives listed. The present section briefly reprises some key findings and addresses the final objective.

This study was conducted for the DUSD (R) based on its expressed concern that CEAT may often be performed poorly or not at all and that the Services may adopt new training systems without adequate justification. No empirical data were gathered during the study to validate or reject this premise.[22] However, the analyses revealed that there are reasons why CEAT may not be performed well in the Services. The *Introduction* of this report suggested four key possibilities for the possible breakdown of the CEAT process in the Services:

- CEAT methods are inadequately defined
- DoD policy guidance is inadequate
- CEAT procedural guidance is inadequate
- Services lack adequate CEAT programs

The next four subsections consider the study's findings as they relate to these four possibilities and identify potential actions that might be taken.

[22] During interviews, several SMEs provided anecdotal evidence supporting the premise.

CEAT Methods Are Not Well Defined

The *CEAT Concepts* and *CEAT Methods* sections of this report sketched the current state of knowledge on conducting CEAT. The analysis revealed that CEAT is not a single method but a family of related methods. The cost analysis part of CEAT is fairly well defined. However, performing the related TEA poses at least two problems: (1) deciding what type of TEA to perform and (2) actually performing the TEA.

This study did not reveal any systematic method for deciding what type of TEA to perform. It follows that it would be useful to provide analysts with a method and guidelines to help them make the most appropriate choice under different conditions.

After one has decided what type of TEA to perform, the next step is to perform the CEAT. In doing so, one may seek guidance from the large body of literature on CEAT and related methods. The trouble is, this information does not exist in an integrated and coherent form in any single document. It follows that it would be helpful to provide analysts with procedural guidance. This is more difficult than it might sound because of the many different possible ways a CEAT can be conducted. There can be no single method to fit every situation. To be truly useful, such guidance must dovetail with a method for deciding what type of CEAT to perform. Another essential element is to provide numerous case studies and examples that can be used by readers.

The literature review revealed that methods of collective training assessment are not fully developed. Conducting CEAT for systems intended to train groups of people is challenging because so little has been done and documented. Orlansky (1994) has observed that the various CEAT methods, concepts, and directives seem mostly to deal with individual training and that it is unclear how well, or whether, they apply to collective or joint training. More R&D is needed to refine methods for (a) collective training assessment and (b) performing CEAT with systems involving collective tasks.

The literature review and interviews with SMEs revealed that (a) development of analytical methods has languished in recent years due to lack of resources, (b) methods are often perceived by users to be difficult to apply and to lack "user friendliness," (c) methods lack validation by comparison of their results with empirical methods, and (d) proponents often find it difficult to convince military decision makers that analytical methods produce valid results. Most of the SMEs who had experience conducting or overseeing research with analytical methods believed that further R&D with them was warranted. The history of this field does not show a consistent pattern of development and improvement. Indeed, the number of different methods and models has often confounded those who (like the author) have ventured into the tangled literature. For more than two decades researchers have worked in this area, yet today no single analytical method is widely accepted and used, and the sources of funding for further developments seem largely to have decided to offer their resources elsewhere. A thoughtful observer might reasonably ask why researchers have so persistently worked in this area while having so little to show for their efforts. The answer is not elusive: analytical methods appear to hold out the promise of providing useful data in situations that preclude empirical methods and at much lower cost. Before abandoning these methods entirely, it would be appropriate to review the developments of the past 20 years, take stock of what lessons have been learned, and decide whether some action by DoD should be taken. Outcomes of this review might be to (1) take no action, which is probably the equivalent of allowing the field of analytical methods to continue its decline; (2) identify and support the validation of one or more promising analytical methods; or (3) support the development of new methods.

DoD Policy Guidance Is Ambiguous

The *CEAT Written Guidance* section of this report reviewed DoD written guidance which could be construed to relate to CEAT and reached several conclusions about key documents. The 5000-series DoD instructions are sufficiently ambiguous about CEAT requirements that it would be reasonable to conclude that they do not require CEAT. They do require COEA above a certain dollar threshold and if a new training system exceeds this threshold, then one might construe the instructions as requiring the training equivalent of a COEA (i.e., CEAT); however this is not explicitly stated. In any case, the dollar threshold is so high that very few training systems would be affected. *DoD Directive 1430.13* requires a Service to conduct CEAT-type analyses but seems to exclude training technologies that do not fit the definition of training device, simulator, or system (e.g., distance learning technologies, training delivery media) for which CEAT might be appropriate. Without a clearly stated requirement, the program manager has little incentive to perform analyses that might slow a program or cost money. If DoD policy-makers want to insure that CEATs are performed under certain conditions, then the policy should be presented more clearly in policy documents.

The Army has published regulations making conduct of CEAT Army policy but the Navy and Air Force have not. If the DoD wants to close this gap, options appear to be to (a) direct the Services to develop Service policy guidance similar to the Army's, (b) clarify policy statements in existing DoD directives, or (c) promulgate a DoD CEAT MIL-STD.

CEAT Procedural Guidance Is Inadequate

There are good academic treatments of CBA (e.g., Sassone & Schaffer, 1978) and the Army has developed several useful guides on the conduct of experimentally-based TEA (see *CEAT Written Guidance*), but there is no comprehensive guide on the conduct of CEAT. Existing procedural guidance is fragmented. It must be concluded that CEAT procedural guidance is inadequate.

The author has concluded that *the complexity of CEAT precludes the development of a cookbook-style "how to" guide for conducting CEAT under all circumstances.* It would be more realistic to assemble a set of CEAT resources. The resources envisioned would consist of three elements:

- *Method Selection*: A set of rules to enable the analyst to determine the most suitable CEAT method based on stage of the WSAP, data requirements (e.g., validity, reliability), cost (of system being analyzed and of conducting the analysis), and analysis requirement and constraints (funds, time, operational limitations, personnel), and other factors CEAT experts consider important.

- *Methods*: Descriptions of the general classes of methods presented in Table 5, with suitable samples of data collection instruments, analyses, and other relevant information.
- *Case Studies*: Examples of completed studies linked to each method that can be used as models.

The first element, Method Selection, does not presently exist and would have to be developed. The second element, Methods, exists in fragmented form but would have to be compiled from existing materials where they exist or created where it does not exist. The third element, Case Studies, could be compiled from existing studies and test reports.

CEAT Programs Differ among the Services

The *CEAT in the Services* section of this report defined the existence of a Service CEAT "program" based on published CEAT requirements, organization, and regular publication of reports. It concluded that, by the definition given, the Army has a CEAT program but that the Navy and Air Force do not. Interviews conducted with SMEs revealed that CEAT is not performed in a routine manner in the Navy and Air Force, although it may occur when a perception of need arises. Anecdotal evidence suggests that that there may be problems in Navy and Air Force CEAT, but the study lacks empirical data to support this notion. Even if such evidence existed, it would be inappropriate for the OSD to tell the Services how to operate their CEAT programs. The OSD can, however, insure that DoD CEAT policy and doctrine are comprehensive and explicit, and leave it to the Services to decide how to implement them within each Service's particular organization and culture.

REFERENCES

Adams, A.V. & Rayhawk, M. (1987). *A review of models of cost and training effectiveness analysis (CTEA), Volume II: Cost analysis.* ARI Research Note 87-59. Washington, DC: Consortium of Washington Area Universities. (AD-A189645)

Adams, A.V. & Rayhawk, M. (1988). *Training effectiveness and cost iterative technique (TECIT), Volume II: Cost effectiveness analysis.* ARI Research Note 88-57. Washington, DC: Consortium of Washington Area Universities. (AD-A197429)

Angier, B.N., Alluisi, E.A., & Horowitz, S.A. (1992). *Simulators and enhanced training.* IDA Paper P-2672. Alexandria, VA: Institute for Defense Analyses.

Barr, R. (1986). *Measures of effectiveness in systems analysis and human factors.* NWC Report No. NWC TP 6740. China Lake, CA: U.S. Navy Naval Weapons Center.

Bessemer, D.W. (1991). *Transfer of SIMNET training in the armor officer basic course.* ARI TR 920. Ft. Knox, KY: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A233198)

Boldovici, J.A. (1987). Measuring transfer in military settings. In S.M Cormier & J.D. Hagman (eds.), *Transfer of learning: Contemporary research and applications.* San Diego, CA: Academic Press.

Boldovici, J.A. (1993). *Considerations in the design of training research.* Unpublished manuscript. Orlando, FL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Boldovici, J.A. (1995, January 3). *Review comments to Henry Simpson re cost effectiveness report.* Unpublished manuscript. Orlando, FL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Boldovici, J.A. & Bessemer, D.W. (1994). *Training research with SIMNET: Lessons learned from simulation networking.* ARI Technical Report 1006. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Boldovici, J.A. & Kraemer, R.E. (1975). *Specifying and measuring unit performance objectives.* Alexandria, VA: Human Resources Research Organization. (AD-A081014)

Braby, R., Henry, J.M., Parrish, W.F, & Swope, W.M. (1975). *A technique for choosing cost-effective instructional delivery systems.* TAEG Report No. 16. Orlando, FL: Training Analysis and Evaluation Group. (AD-012859)

Braby, R., Morris, C.L., Micheli, G.S., & Okraski, H.C. (1972). *Staff study on cost and training effectiveness of proposed training systems.* TAEG Report 1. Orlando, FL: Training Analysis and Evaluation Group.

Burnside, B.L. (1990). *Assessing the capabilities of training simulations: A method and simulation network (SIMNET) application.* ARI Research Report 1565. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand McNally.

Clinton, William J. (1994). *Budget of the united states government, fiscal year 1995.* Washington, DC: U.S. Government Printing Office.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45(12)*, 1304-1312.

Cohen, J. (1994, December). The earth is round ($p < .05$). *American Psychologist.. 49(12)*, 997-1003.

Companion, M. (1990). *ASTAR operational evaluation final report, Volume II: Test reports.* IST-TR-90-08. Orlando, FL: Institute for Simulation and Training.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Deitchman, S.J. (1990). *Further explorations in estimating the military value of training.* IDA Paper P-2317. Alexandria, VA: Institute for Defense Analyses.

Deitchman, S.J. (1993). *Quantifying the military value of training for system and force acquisition decisions: An appreciation of the state of the art.* IDA Paper P-2881. Alexandria, VA: Institute for Defense Analyses. (AD-A274753)

Department of the Army (1980). *TRADOC Training effectiveness analysis handbook* (draft). White Sands Missile Range, NM: TRADOC Systems Analysis Activity.

Department of the Army (1981). *Army regulation 5-5: Army studies and analyses.* Washington, D.C.: Department of the Army.

Department of the Army (1985). *TRADOC pamphlet 11-8: Studies and analysis handbook.* Ft. Monroe, VA: U.S. Army Training and Doctrine Command.

Department of the Army (1986). *Methodology for abbreviated analyses.* Ft. Gordon, GA: Directorate of Combat Developments, USASC&FG.

Department of the Army (1988a). *Army regulation 70-1: Systems acquisition policy and procedures.* Washington, D.C.: Author.

Department of the Army (1988b). *TRADOC regulation 350-7: Systems approach to training.* Ft. Monroe, VA: U.S. Army Training and Doctrine Command.

Department of the Army (1993). *TRADOC regulation 350-32: The TRADOC training effectiveness analysis (TEA) system* (draft). Ft. Monroe, VA: U.S. Army Training and Doctrine Command.

Department of Defense (1986). *DoD directive1430.13: Training simulators and devices.* Washington, D.C.: Author.

Department of Defense (1990). *MIL-STD-1379D: Military training programs.* Washington, D.C.: Author.

Department of Defense (1991a). *DoD directive 5000.1: Defense acquisition.* Washington, D.C.: Author.

Department of Defense (1991b). *DoD instruction 5000.2: Defense acquisition management policies and procedures.* Washington, D.C.: Author.

Department of Defense (1991c). *DoD instruction 5000.2M: Defense acquisition management documentation and reports.* Washington, D.C.: Author.

Department of Defense (1994). *Military manpower training report: FY 1995.* Washington, D.C.: Author.

Department of the Navy (1985). *OPNAV instruction 5311.7: Determining manpower, personnel and training (MPT) requirements for Navy acquisitions.* Washington, D.C.: Chief of Naval Operations (OP-111).

Department of the Navy (1987). *The navy program manager's guide to early MPT planning.* Washington, D.C.: Chief of Naval Operations (OP-111) HARDMAN Development Office.

Derrick, W.L. & Davis, M.S. (1993). *Cost-effectiveness analysis of the C-130 aircrew training system.* AL-TR-1992-0173. Arlington, VA: Systems Research and Applications Corporation. (AD-B171592)

Djang, P.A., Butler, W.G., Laferriere, R.R., & Hughes, C.R. (1993). *Training mix model.* TRAC-WSMR-TEA-93-035. White Sands Missile Range, NM: TRADOC Analysis Center

Ellis, J., Knirk, F., Taylor, B., & McDonald, B. (1987). *The course evaluation system.* NPRDC TR 87-19. San Diego, CA: Navy Personnel Research and Development Center.

Faust, D.G., Swezey, R.W., & Unger, K.W. (1984). *Field application of TRAINVICE: A study of four models designed to predict training device transfer of training potential.* McLean, VA: Science Applications, Inc.

Festinger, L, and Katz, D. (1953). *Research methods in the behavioral sciences.* New York: Dryden Press.

Fishburne, R.P. & Rolnick, S.J. (1985). *Guidelines for conducting a training effectiveness evaluation (TEE), Volume I: TEE evaluator's handbook.* ARI Research Product 85-14. Buffalo, NY: Calspan Corporation. (AD-A170388)

Frederickson, E.W. (1981). An enriched user oriented cost and training effectiveness analysis methodology. *Proceedings of the 23rd. Annual Conference of the Military Testing Association, 1*, 429-438. (AD-P001317)

Frost & Sullivan (1994). *U. S. military trainers and simulators markets (volume 1)* New York, NY: Frost and Sullivan Market Intelligence.

Gagne, R.M., Foster, H., & Crowley, M.E. (1948). The measurement of transfer of training. *Psychological Bulletin, 45(2).* 97-130.

Gibbons, S. & Franchi, J. (1990). *Analytic investigation: A comparison of ASTAR and other device effectiveness techniques.* IST-TR-90-09. Orlando, FL: University of Central Florida/Institute for Simulation and Training.

Gibson, R.S. & Orlansky, J. (1986). *Performance measures for evaluating the effectiveness of maintenance training.* IDA Paper P-1922. Alexandria, VA: Institute for Defense Analyses.

Goldberg, I. (1988). *Training effectiveness and cost iterative technique (TECIT), Volume I: Training effectiveness analysis.* ARI Research Note 88-35. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A196727)

Goldberg, I. & Khattri, N. (1987). *A review of models of cost and training effectiveness analysis (CTEA), Volume I: Training effectiveness analysis.* ARI Research Note 87-58. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A189198)

Gorman, P.F. (1990). *The military value of training.* IDA Paper P-2515. Alexandria, VA: Institute for Defense Analyses.

Hall, E.R., Kincaid, J.P., Braby, R., Kiernan-Kostic, D., Muller, J.K., & Walker, B. (1994). *Cost and training effectiveness analysis: II. Procedural guidance.* IST-PR-94-19. Orlando, FL: University of Central Florida/Institute for Simulation and Training.

Hall, E.R., Kincaid, J.P., Muller, J.K., & Kiernan-Kostic, D. (1994). *Cost and training effectiveness analysis: I. Acquisition of data.* IST-PR-94-18. Orlando, FL: University of Central Florida/Institute for Simulation and Training.

Hall, E.R., Rankin, W.C., & Aagard, J.A. (1976). *Training effectiveness assessment, Volume II: Problems, concepts, and evaluation alternatives.* TAEG Report No. 39. Orlando, FL: Training Analysis and Evaluation Group.

Hammon, C.P. & Horowitz, S.A. (1987). *Relating personnel and training resources to unit performance.* IDA Paper P-2023. Alexandria, VA: Institute for Defense Analyses.

Hoffman, R.G. & Morrison, J.E. (1992). *Methods for determining resource and proficiency tradeoffs among alternative tank gunnery training methods.* ARI Research Product 92-03. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A250867)

Jeantheau, G.G. (1971). *Handbook for training systems evaluation.* NAVEDTRACEN 66-C-0113-2. Darien, CT: Dunlap & Associates. (AD-733962)

Klein, G.A., Johns, P., Perez, R., & Mirabella, A. (1985). *Comparison-based prediction of cost and effectiveness of training devices: A guidebook.* ARI Research Product 85-29. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A170941)

Knapp, M.I. & Orlansky, J. (1983). *A cost element structure for defense training.* IDA Paper P-1709. Alexandria, VA: Institute for Defense Analyses.

Knerr, C.M., Nadler, L.B., & Dowell, S.K. (1984). *Training transfer and effectiveness models.* HumRRO FR-DRD(VA)-84-1. Alexandria, VA: Human Resources Research Organization.

Larsen, J.Y., Rolnick, S.J., & Fishburne, R.P. (1985). *Guidelines for conducting a training effectiveness evaluation (TEE), Volume II: Data collector's manual* ARI Research Product 85-15. Buffalo, NY: Calspan Corporation. (AD-A170389)

Martin, M.F., Rose, A.M., & Wheaton, G.R. (1988). *Applications of ASTAR in training system acquisitions.* AIR-49901-TR1-01/88. Washington, DC: American Institutes for Research.

Matlick, R.K., Berger, D.C., Knerr, C.M., & Chiorini, J.R. (1980). *Cost and training effectiveness analysis in the army life cycle systems management model.* ARI TR 503. Ft. Bliss, TX: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A109198)

Matlick, R.K., Berger, D.C., & Rosen, M.H. (1980). *Cost and training effectiveness analysis (CTEA) performance guide*. Research Product 81-1. Ft. Bliss, TX: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A101985)

McFann, H. (1983). *Candidate application analyses for the ARI/NTC data base*. Monterey, CA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McFann, H. (1990). *Methods for measurement of army unit performance at combat training centers*. Orlando, FL: Proceedings of the Cost-Effectiveness Analysis of Training Systems Workshop of the 11th. Interservice/Industry Training Systems Conference, December 1990.

McMichael, J.S. (1985). *Cost-benefit analysis in military manpower and training research and development: Current practices.* Symposium on the Military Value and Cost-Effectiveness of Training, 7-9 January 1985. Brussels, Belgium: NATO

Micheli, G. (1994, October 28). *Personal communication.*

Morrison, J.E. & Hoffman, R.G. (1992). *A user's introduction to determining cost-effective tradeoffs among tank gunnery training methods.* ARI Research Note 92-29. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A250-029)

Muckler, F.A. & Finley, D.L. (1994a). *Applying training system estimation models to army training, Volume I: Analysis of the literature.* ARL-TR-463. Aberdeen Proving Grounds, MD: U.S. Army Research Laboratory.

Muckler, F.A. & Finley, D.L. (1994b). *Applying training system estimation models to army training, Volume II: An annotated bibliography 1970-1990.* ARL-TR-463. Aberdeen Proving Grounds, MD: U.S. Army Research Laboratory.

O'Neil, H.F., Baker, E.L., & Kazlauskas, E.J. (1990). Assessment of team performance. In. R.W. Swezey & E. Salas (eds.), *Teams: Their training and performance.* Norwood. NJ: Ablex.

Orlansky, J. (1985). *Panel on the defence applications of operational research*. DA/A/DR (85)167. Proceedings of the Symposium on the Military Value and Cost-Effectiveness of Training, 7-10 January 1985. NATO, Brussels, Belgium. (AD-B-093-505)

Orlansky, J. (1989). *Executive Summary.* AC/243 (Panel 7/RSG.15)D/4 on the Military Value and Cost-Effectiveness of Training. NATO, Brussels, Belgium.

Orlansky, J. (1994, November 1). *Personal communication.*

Osgood, C.E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review, 56,* 132-143.

Paris, H.L. (1990). *Training effectiveness (TEA) contributions of the army*. Orlando, FL: Proceedings of the Cost-Effectiveness Analysis of Training Systems Workshop of the 11th. Interservice/Industry Training Systems Conference, December 1990.

Pfeiffer, M.G. & Browning, R.F. (1984). *Field evaluations of aviation trainers.* NTSC TR-157. Orlando, FL: Naval Training Systems Center. (AD-B083584).

Pfeiffer, M.G., Evans, R.M, & Ford, L.H. (1985). *Modeling field evaluations of aviation trainers.* TAEG TN 1-85. Orlando, FL: Training Analysis and Evaluation Group.

Pfeiffer, M.G. & Horey, J.D. (1988). *Analytic approaches to forecasting and evaluating training effectiveness.* NTSC TR-88-027. Orlando, FL: Naval Training Systems Center. (AD-B129158)

Povenmire, H.K. & Roscoe, S.N. (1972). Incremental transfer effectiveness of a ground based aviation trainer. *Human Factors, 15*, 534-542

Powers, T.R., McCluskey, M.R., Haggard, D.F., Boycan, G.G., & Steinheiser, F. (1974). *Determination of the contribution of live firing to weapons proficiency.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rankin, W. & Swope, W. (1990). *Naval training systems center techniques and lessons learned.* Orlando, FL: Proceedings of the Cost-Effectiveness Analysis of Training Systems Workshop of the 11th. Interservice/Industry Training Systems Conference, December 1990.

Resource Consultants, Inc. (1992). *Measures of crew/team training effectiveness for cost/training effectiveness analysis: Training effectiveness catalog system (TECATS) database.* Orlando, FL: Author.

Resource Consultants, Inc. (1993). *Measures of crew/team training effectiveness for cost/training effectiveness analysis.* Orlando, FL: Author.

Rolnick, S.J., Fishburne, R.P., & Nawrocki, L.H. (1985). *Guidelines for conducting a training effectiveness evaluation (TEE), Volume III: User's guide for revising training program deficiencies.* ARI Research Product 85-16. Buffalo, NY: Calspan Corporation. (AD-A170621)

Roscoe, S.N. (1971). Incremental Transfer Effectiveness. *Human Factors, 13*, 561-567.

Roscoe, S.N. (1972). A little more on incremental transfer effectiveness. *Human Factors, 14*, 363-364

Rose, A.M. & Martin, M.F. (1988). *Forecasting training system effectiveness: DEFT final report.* TCN TR-87-430. Washington, DC: American Institutes for Research.

Rose, A.M., Wheaton, G.R., & Yates, L.G. (1985a). *Forecasting device effectiveness: I. Issues.* ARI TR 680. Washington, DC: American Institutes for Research. (AD-A159576)

Rose, A.M., Wheaton, G.R., & Yates, L.G. (1985b). *Forecasting device effectiveness: II.. Procedures.* ARI Research Product 85-25. Washington, DC: American Institutes for Research. (AD-159955)

Rose, A.M. & Martin, A.W., & Yates, L.G. (1985). *Forecasting device effectiveness: III. Analytic assessment of device effectiveness forecasting technique.* ARI TR 681. Washington, DC: American Institutes for Research. (AD-160029)

Rosen, M.H. , Berger, D.C., & Matlick, R.K. (1985). *A review of models for cost and training effectiveness analysis.* ARI Research Note 85-34. Ft. Benning, GA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A158041)

Rupinski, T.E. & Stoloff, P.H. (1990). *An evaluation of navy video teletraining (VTT).* CRM 90-36. Alexandria, VA: Center for Naval Analyses. (AD-A239180)

Salas, E., Dickinson, T.L. Converse, S.A., & Tannenbaum, S.I. (1992). Toward an understanding of team performance and training. In R.W. Swezey & E. Salas (eds.), *Teams: Their training and performance.* Norwood. NJ: Ablex.

Sassone, P.G. (1985). *Training extension course research: Review of the literature on cost and training effectiveness.* ARI Research Note 85-74. Ft. Benning, GA: Litton Systems, Inc. (AD-A160632)

Sassone, P.G. & Schaffer, W.A. (1978). *Cost-benefit analysis: A handbook.* New York, NY: Academic Press.

Sheppard, D.J., Hassen, J.E., Hodak, G.W., Swope, W.M., & Denton, C.F. (1991). *An evaluation and cost analysis of video teletraining applications for the naval reserve training program.* Special Report 91-002. Orlando, FL: Naval Training Systems Center.

Sheppard, D.J. & Hodak, G.W. (1991). *Evaluation and travel cost avoidance analysis of proposed video teletraining stations.* Technical Report 91-005. Orlando, FL: Naval Training Systems Center.

Singer, M.J. (1994, November 3). *Personal communication.*

Singer, M.J. (1993). *The optimization of simulation-based training systems: A review of evaluations and validation of rule bases.* ARI Research Report 1653. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A278149)

Smith, B.R. (1994). *Cost and training effectiveness analysis: Work unit efforts & studies and analysis efforts.* San Diego, CA: Manpower and Training Research Information System.

Solomon, H. (1986). *Economic issues in cost-effectiveness analyses of military skill training.* IDA Paper P-1897. Alexandria, VA: Institute for Defense Analyses. (AD-A171106)

Sticha, P.J. (1994, October 31). *Personal communication.*

Sticha, P.J. (1989). Normative and descriptive models for training-system design. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, & L. Van Breda (eds.), *Applications of human performance models to system design.* New York, NY: Plenum.

Sticha, P.J., Blacksten, H.R., Knerr, C.M., Morrison, J.E., & Cross, K.D. (1986). *Optimization of simulation based training systems, Volume II: Summary of the state of the art.* HumRRO FR-PRD-86-13. Alexandria, VA: Human Resources Research Organization.

Sticha, P.J., Blacksten, H.R., Buede, D.M., & Cross, K.D. (1986). *Optimization of simulation based training systems, Volume III: Model description.* HumRRO FR-PRD-86-13. Alexandria, VA: Human Resources Research Organization.

Sticha, P.J., Blacksten, H.R., Buede, D.M., Singer, M.J., Gilligan, E.L., Mumaw, R.J., & Morrison, J.E. (1990). *Optimization of simulation-based training systems: Model description, implementation, and evaluation.* ARI TR 896. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sticha, P.J., Singer, M.J., Blacksten, H.R., Morrison, J.E., & Cross, K.D. (1990). *Research and methods for simulation design: State of the art.* HumRRO FR-PRD-88-27. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Swezey, R.W. & Salas, E. (1992). Guidelines for use in team training development. In. R.W. Swezey & E. Salas (eds.), *Teams: Their training and performance.* Norwood. NJ: Ablex.

Swope, W.M. (1976). *A primer on economic analysis for naval training systems.* TAEG Report No. 31. Orlando, FL: Training Analysis and Evaluation Group.

Taylor, B.E., Ellis, J.A., & Baldwin, R.L. (1987). *Current status of navy classroom training: A review of 100 navy courses with recommendations for the future.* NPRDC TR 88-11. San Diego, CA: Navy Personnel Research and Development Center.

TRADOC (1993a). *The TRADOC DCST's AR 5-5 training studies program.* Ft. Monroe, VA: Author.

TRADOC (1993b). *FY94 TRADOC Study Program Study Descriptions.* Ft. Monroe, VA: Author.

Tufano, D.R. & Evans, R.A. (1982). *The prediction of training device effectiveness: A review of army models.* ARI TR 613. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A146937)

Turnage, J.J., Houser, T.L., & Hofmann, D.A. (1990). *Assessment of performance measurement methodologies for collective military training.* ARI Research Note 90-126. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A227971)

Tuttle, T.C. & Weaver, C.N. (1986). *Methodology for generating efficiency and effectiveness measures: A guide for air force measurement facilitators.* AFHRL Technical Paper 86-36. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

U.S. Army Armor School (1989). *SIMNET users' guide*. Fort Knox, KY: Author.

Waag, W.L., Pierce, B.J., & Fessler, S. (1987). *Performance measurement requirements for tactical aircrew training.* AFHRL-TR-86-62. Williams Air Force Base, TX: Air Force Human Resources Laboratory.

Witmer, B.G. (1991). *VALTRAIN: A conceptual model for estimating the value of training devices/systems.* Working Paper STRICOM, Orlando FU WP-91-1. Orlando, FL: U.S. Army Simulation, Training and Instrumentation Command.

Zimmerman, W., Butler, R, Gray, V, Rosenberg, L., & Risser, D. (1984). *Evaluation of the HARDMAN (hardware vs. manpower) comparability methodology.* ARI TR-646. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A162847)

APPENDIX

ABBREVIATIONS AND ACRONYMS

APPENDIX: ABBREVIATIONS AND ACRONYMS

AFHRL	Air Force Human Resources Laboratory
APS	Analytic Profile System
ARI	U.S. Army Research Institute for the Behavioral and Social Sciences
ASTAR	Automated Simulator Test and Assessment Routine
ATM	Analogous Task Method
BDM/CARAF	BDM service company Combined Arms Research and Analysis Facility
CA	Cost Analysis or Conjoint Analysis
CBA	Cost-Benefit Analysis
CBP	Comparison-Based Prediction
CEA	Cost-Effectiveness Analysis
CEAT	Cost-Effectiveness Analysis of Training
CHRT	Coordinated Human Resources Technology
CK	Checklist
CNA	Center for Naval Analyses
CNET	Chief of Naval Education and Training
COEA	Cost and Operational Effectiveness Analysis
CTEA	Cost and Training Effectiveness Analysis
CTER	Cumulative Transfer Effectiveness Ratio
DEFT	Device Effectiveness Forecasting Technique
DEI	Display Evaluation Index
DHQ	Device Handling Qualities
DMDC	Defense Manpower Data Center
DoD	Department of Defense
DRIMS	Diagnostic Rifle Marksmanship Simulators
DUSD (R)	Deputy Under Secretary of Defense for Readiness
FA	Fidelity Analysis
FORTE	Forecasting Training Effectiveness
HARDMAN	Hardware Manpower Integration
HUMRRO	Human Resources Research Organization
IDA	Institute for Defense Analyses
IQI	Instructional Quality Inventory
MATRIS	Manpower and Training Research Information System
MAU	Multiattribute Utility Analysis
MDSA	Multidimensional Scaling Analysis
METT-T	Mission, Enemy forces, Troops friendly, Terrain control, Time
MIL-STD	Military Standard
MMM	Multitrait-Multimethod Matrix
MODIA	Method of Designing Instructional Alternatives
MOE	Measure of Effectiveness
MPT	Manpower, Personnel, and Training

NAWC	Naval Air Warfare Center
NAWCTSD	Naval Air Warfare Center Training System Division
NTDC	Naval Training Device Center
NTEC	Naval Training Equipment Center
NTSC	Naval Training Systems Center
NPRDC	Navy Personnel Research and Development Center
ORSA	Operations Research/Systems Analysis (also: Operations Research Society of America)
OSBATS	Optimization of Simulation-Based Training Systems
OSD	Office of the Secretary of Defense
P/C/D	Training Programs, Courses, and Devices
POI	Program of Instruction
R&D	Research and Development
RCI	Resource Consultants, Incorporated
SAT	Systems Approach to Training
SME	Subject-Matter Expert
SOMA	System Operability Measurement Algorithm
ST	Simulated Transfer
STC	Simulator Training Capability
STRICOM	U.S. Army Simulation, Training and Instrumentation Command
TADSS	Training Aids, Devices, Simulators, and Simulations
TAEG	Training Analysis and Evaluation Group
TCA	Task Commonality Analysis or Training Consonance Analysis
TD	Training Device
TD/S	Training Device/Simulator
TDDA	Training Developer's Decision Aid
TDDSS	Training Development Decision Support System
TEA	Training Effectiveness Analysis
TECEP	Training Effectiveness and Cost-Effectiveness Prediction
TECIT	Training Effectiveness and Cost Iterative Technique
TEEM	Training Efficiency Estimation Model
TER	Transfer Effectiveness Ratio
TIM	Training Interlock Measure
TRAC	TRADOC Analysis Center
TRADOC	Training and Doctrine Command
TRAINVICE	Training Device Effectiveness Model
TRAM	Training Analysis Model
TRAMOD	Training Requirements Analysis Model
VALTRAIN	Value of Training Model
VTT	Video Teletraining
WSAP	Weapon System Acquisition Process