DTIC FILE COPY

ARI Research Note 87-58

AD-A189 198

A REVIEW OF MODELS OF COST AND TRAINING EFFECTIVENESS ANALYSIS (CTEA)
Volume I: Training Effectiveness Analysis

Isadore Goldberg and Nidhi Khattri
The Consortium of Washington Area Universities

for

Contracting Officer's Representative
John J. Kessler

Training and Simulation Technical Area
Robert J. Seidel

TRAINING RESEARCH LABORATORY
Jack H. Hiller, Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

October 1987

Approved for public release; distribution unlimited.

87 12 14 041

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS None
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARI Research Note 87-58	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a NAME OF PERFORMING ORGANIZATION Consortium of Washington Area Universities 6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) 1717 Massachusetts Avenue, NW, # 101 Washington, DC 20036	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Same as in 7a. 8b. OFFICE SYMBOL (If applicable) PERI	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-82-C-0383
8c. ADDRESS (City, State, and ZIP Code) Same as in 7b.	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 6.37.44.A PROJECT NO. 2Q263744A795 TASK NO 3.4.1. WORK UNIT ACCESSION NO. 3.4.1.R.1

11 TITLE (Include Security Classification)
A Review of Models of Cost and Training Effectiveness Analysis, Volume I: Training Effectiveness Analysis

12. PERSONAL AUTHOR(S)
Isadore Goldberg and Nidhi Khattri

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM Aug 85 TO Aug 86	October 1987	160

16. SUPPLEMENTARY NOTATION John J. Kessler, contracting officer's representative. Volume II. Cost Analysis by A.V. Adams & M.L. Rayhawk, George Washington University, Washington, D.C.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Models Training Effectiveness Analysis (TEA) Models Training Devices (OVER)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research note reviews the formal predictive and prescriptive models of training effectiveness analysis and related Army guidance. Most of the models are concerned with the analytic formation of entire training programs early in the weapons systems acquisition process. Some models also include manpower, personnel, and human factors considerations. A number of models focus on the formulation of training devices and simulators. Computer Based Instruction (CBI) has not often been included in early formation of training programs for new weapons systems. Possible reasons for this finding are discussed. Most of the training program models also include a cost model, but the cost models are not necessarily adequate. The models concerned with training devices and simulators do not have associated cost models.

The lessons learned included the following: 1) there are many useful models available for the formulation of training programs early in the weapons system acquisition process: 2) the validity of these models needs to be tested, recommendations are given for comparative validity studies; 3) models for training devices and simulators appear to need (OVER)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code) / 22c OFFICE SYMBOL

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARI RESEARCH NOTE 87-58

18. Subject Terms (continued)

Computer Based Instruction (CBI)
Cost and Training Effectiveness Analysis (CTEA)

20. Abstract (continued)

further development, related cost models, and validation; 4) further attention needs to be given to formal CTEA models for advanced phases of the weapons system acquisition process, and for non-system training.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

The Consortium of Washington Area Universities

Technical review by

Douglas Dressel

Douglas A. Ramsay

FOREWORD

One of the missions of the Army Research Institute for the Behavioral and Social Sciences (ARI) is to conduct research and development in individual training methods. The Army has established the Training Technology Field Activities (TTFA) to determine how and where training technology can be used to provide cost-effective military training . TTFAs are located at several Army schools and are organized as partnerships consisting of ARI, the Training and Doctrine Command (TRADOC), and the particular school. This report is part of ARI's support of the TTFAs.

The task to develop a theoretical model which would enable TTFAs to systematically compare alternative ways of training on combined cost and effectiveness variables was contracted to the Consortium of Universities of the Washington Metropolitan Area and monitored by ARI. The contract with the Consortium is one of a series let by ARI intended to bring the expertise of academia to bear on the Army's personnel research and also support the Defense Department's effort to strengthen the university research base.

The products of this contract are four reports, two which review pertinent literature and two which treat the development of the theoretical model. Anyone concerned with cost and training effectiveness analysis or the requirements of model development will find these reports comprehensive and thorough.

ACKNOWLEDGMENTS

This report was prepared with support from the Consortium of Washington Area Universities and the Army Research Institute for the Behavioral and Social Sciences. We are indebted to a number of individuals who have made important contributions to the study. We wish to especially thank the Army Research Institute Advisory Committee, including D. Bruce Bell, Stanley F. Bolin, Hyder Lakhani, Donald Haggard, Harold Wagner and John J. Kessler, who offered advice and counsel throughout the study and who served as a sounding board for our ideas. A special note of appreciation is offered to John J. Kessler who served as the Contracting Officer's Representative. Dr. Kessler provided valuable assistance throughout the effort and made many useful suggestions for improving the final report.

During the course of the study, several individuals at the Ft. Knox Armor Center provided us with access to specialized literature and insights into the conduct of training using computer-assisted instruction, training devices and simulators. These insights have materially enriched the study. Among others who contributed in this respect, we want to especially thank Donald F. Haggard, Donald M. Kristiansen and John A. Boldovici. From our own staff we wish to thank Twannah Ellington for typing and production of this report. As is customary, the views expressed in this study are our own and do not necessarily reflect the official views of the U.S. Army. We naturally assume responsibility for any errors or omissions.

Isadore Goldberg
Nidhi Khattri

A REVIEW OF MODELS OF COST AND TRAINING EFFECTIVENESS ANALYSIS, VOLUME I: TRAINING EFFECTIVENESS ANALYSIS

EXECUTIVE SUMMARY

Requirement:

To identify cost and training effectiveness analysis (CTEA) models which might be adapted to the needs of the Army for improving its training development system, particularly in areas involved with use of new training technology. The focus of this report is on models that deal with training effectiveness analyses. The companion report, Volume II: Cost Analysis, examines models concerned with measures of training costs.

Procedure:

A body of literature assembled by the Army Research Institute and literature already in possession of the authors was searched. This literature enabled the devlopment of a keyword search of computerized data bases which included the Defense Technical Information Center (DTIC), the National Technical Information Service (NTIS), the Educational Resources Information Center (ERIC), and PsychScan. References were also obtained through contact with several training and cost experts and major consulting firms.

Findings:

The report reviews formal predictive and prescriptive models of training effectiveness analysis and related Army guidance. Lessons learned include the following: (1) There are many useful models available for the formulation of training programs early in the weapons system acquisition process. (2) The validity of these models needs to be tested. Recommendations are given for comparative validity studies. (3) Models for training devices and simulators appear to need further development, related cost models, and validation. Further attention needs to be given to formal CTEA models for advanced phases of the weapons systems acquisition process and for non-system training.

Utilization of Findings:

This literature review will be valuable to researchers who need a grounding in CTEA models. It will also be of interest to decision-makers who wish to know something of the internal workings of a model and where the trade-offs reside. The authors are using the contents of this report to develop a CTEA model called Training and Cost Iterative Technique (TECIT) which will be described in subsequent publications.

Table of Contents

TABLES

FIGURES

Chapter 1

INTRODUCTION

PURPOSE AND SCOPE OF THE REPORT

It is now almost twenty years* since the Army recognized that fielding of weapon systems (WS) required early consideration of the impact of training costs and effectiveness. Since that time, many efforts have been undertaken to develop formal models appropriate for a variety of training problems and contexts.

The purpose of this report is to review the Cost and Training Effectiveness Analysis (CTEA) models, especially "predictive" models, appropriate to the planning of training, and the selection of media, methods, and training devices early in the Weapon System Acquisition Process (WSAP). The review assesses the "state-of-the-art" in the field, the rationale underlying the various models, and makes recommendations for further testing and development of the models. This report emphasizes training effectiveness while a companion report emphasizes cost analysis.

The two reviews are to be used as a point of departure for a second report that will propose a generic model of CTEA, one that can be applied to a variety of training applications in Army settings. The resulting analysis and model development is oriented toward the needs of the Army Training Technology Field Activities (TTFAs), a new unit

*The first version of DA Pamphlet 11-25, Life Cycle Systems Management Model for Army System was published in October 1968, but the problem was being considered for several years prior to this.

within TRADOC charged with the improvement of training, as well as other interested Army and DOD organizations.

This report builds on earlier reviews of CTEA models by Matlick et al. (1980), Rosen et al. (1981), and Knerr et al. (1984). Although each of these reports had somewhat different emphases, they more than adequately reviewed issues related to the following:

1. DOD and Army guidance related to CTEA models.
2. Input information required to use CTEA models.
3. The processes, algorithms, and metrics of CTEA models.
4. Prescriptive CTEA models.
5. Empirical models in CTEA.

Transfer of training models reviewed by Tufano and Evans (1984) and Knerr et al. (1984) also helped focus this review. Consequently, this report emphasizes new developments and issues attendant to the further development of the predictive models.

DEFINITIONS OF CTEA MODELS

Emphasis in these reports is on formal CTEA models. Formal models are defined as those that employ algorithms, decision rules and mathematical formulas in structuring measures of effectiveness and costs. Other documents that provide guidance and informal models are also included where appropriate, when no formal model exists.

Several reviewers (Knerr et al., 1984; Matlick et al., 1981) referred to planning models as predictive and prescriptive models to emphasize differences in the variables they attempt to account for early in the WSAP. They define predictive models as those which predict the extent of effectiveness if training follows a certain course. Prescriptive models are defined as those that outline the course of instruction. Empirical models are defined as those used in evaluating training after it is fielded.

Our definitions are somewhat different. We consider a model predictive when it generates quantitative measures of course performance, transfer, course time estimates, student throughput etc. We define a model as prescriptive when it concerns itself with outlining the media, methods and delivery systems of instruction.

Knerr et al. note the lack of validation criteria for the models. For this reason this report points out how empirical data, obtained after a WS and training program has been fielded, can be formulated as criteria for the planning CTEAs. (See Chapter 8.)

CTEA as a specific set of methodologies has been defined primarily in relation to newly developing weapon systems. System training applies to training or training devices developed specifically for the new WS. The task entails obtaining reliable and valid estimates of training requirements and resources and examining trade-offs among hardware, human factors, manpower, personnel and training subsystems. The CTEAs in the early phases of WS development by definition are planning models, as empirical data for the WS and training program will not be available until fielded.

Non-system training refers to training or training devices which are to be used in conjunction with two or more WS or for general military training, such as military courtesy. Only two CTEA efforts were identified in the literature related to non-system training (Rosen et al., 1981 and Kraft and Farr, 1984), suggesting that further work is needed in this area.

Dawdy and Hawley (1982) conceive of CTEA system methods as continua from the planning through the fielding phases, as follows:

> The specific nature of CTEA often is dependent upon the state of development of the materiel system under study. For conceptual materiel systems, the lack of performance data requires that CTEA be used to forecast training resource requirements and to indicate training-related issues that may require special examination during later field testing. As prototypes of the materiel system become available, CTEA involves updating and validating cost and resource impact projections and empirical investigations of training program effectiveness. Following the field deployment of a materiel system, the emphasis of CTEA shifts to the cost-effectiveness of: (1) training "fixes" designed to address recognized training deficiencies, or (2) training modifications designed to meet an altered threat scenario or to accommodate evolutionary hardware modifications.

The reader will note many other points of view represented in the CTEA models reviewed in this report. For example, many of the models are general purpose models for training prescription or transfer of training prediction. Many others were developed for a specific weapon system. One general model combines prescription and prediction, another emphasizes the hardware and human performance trade-offs, a third emphasizes training resource management, and a fourth computer based instruction. This review integrates the literature, examines methodological issues and makes recommendations for further testing, improvement and integration of the models.

LITERATURE SEARCH

An interdisciplinary team of psychologists and economists reviewed the training effectiveness and cost analysis portions of the models. Literature provided by the Army Research Institute (ARI) and literature in the possession of the authors led to the identification of other sources and served as the means for developing a keyword search of computerized data bases spanning civilian and military literature. The data bases searched included the Defense Technical Information Center (DTIC), the National Technical Information Service (NTIS), the Educational Resources Information Center (ERIC), and PsychScan. The search also included references obtained through contact with several training and cost experts and consulting firms.

The DTIC search yielded some 350 abstracts. ERIC and PsychScan identified approximately 250 references, while NTIS produced an additional 300 references. An examination of these abstracts for their relevance to the task at hand led to a narrowing of the number of studies actually reviewed. A bibliography of the literature reviewed is included in the appendix.

ORGANIZATION OF THE REPORT

The report is organized so that chapters 2 through 6 review the models and chapters 7 and 8 review issues and approaches to resolving them.

In brief, the chapters are as follows:

Chapter 2 - The Context of Predictive CTEA: Training Effectiveness Analysis (TEA) and Instructional Systems Development (ISD) Processes in the Weapon System Acquisition Process (WSAP).

Reviews the guidance documents and responsibilities related to CTEA and other methods relating training and training devices to the WSAP. Discusses system and non-system training.

Chapter 3 - Cost and Training Effectiveness in Army Materiel Acquisition.

Reviews a number of documents and specialized models that relate integrated concepts of hardware, manpower, personnel and training.

Chapter 4 - Predictive Cost and Training Effectiveness Models.

Predictive models, as defined, are general models concerned with the quantitative variables of training programs, such as performance, training time, instructor time, and transfer of training. This review updates models not included in detail in earlier reports such as the Training Development Decision Support ystem (TDDSS), Job Skills Education Program CTEA for Computer Based Instruction, Comparison Based Prediction for training devices and the Device Effectiveness Forecasting Technique (DEFT), a new development in predicting transfer of training. TDDSS's contributions in defining subject matter expert qualifications and its method of weighting effectiveness measures are discussed. The discussion of DEFT is integrated with earlier TRAINVICE models. Applications, reliability, validity, performance and time estimating are discussed and a number of recommendations made for further development and validation.

Chapter 5 - Prescriptive Cost and Training Effectiveness Models.

Defined as those that are used to design the media, methods and delivery systems of training, devices or simulators, most of the models employ "figures of merit" based on scoring learning algorithms and media method lists, but do not consider performance or time estimating on the effectiveness side of the equation. These models are reviewed only in brief, deferring to detailed reviews cited earlier vis-a-vis details of input requirements, algorithms, formulas and the like. Each model is a general purpose model, but all purport to serve very similar purposes.

Chapter 6 - Computer Based Instruction (CBI) and Computerized Adaptive Testing (CAT).

CBI is treated as a promising technology for non-system applications. The review notes that only one predictive CTEA model has been proposed so far (and is still in development) specific to CBI; examines the potential suitability of the available predictive and prescriptive models for CBI; reviews the empirical literature; and recommends further developments related to CBI and non-system training. CAT is noted as an emerging technology that may in the future hold promise for reducing achievement testing time with no sacrifice in test reliability and validity.

Chapter 7 - Additional Issues in Training Effectiveness Analysis in CTEA Models.

This chapter starts with a review of findings and observations regarding the predictive and prescriptive models discussed in chapters 4, 5 and 6; abstracts a list of variables and constructs from this literature; reviews and critiques the models in relation to the variables and constructs identified (applications and types of training; performance and time estimates, student throughput and instructional management); considers issues of motivation and responsiveness of the models to policy changes; summarizes the contributions of the models and suggests areas for further development.

Chapter 8 - Methodological Issues in Training Effectiveness Analysis in CTEA.

This chapter focuses upon methodological issues in training effectiveness analysis for CTEA. The discussion of information and data sources points toward an integrated concept of information requirements, and use of subject matter experts in relation to analytic methods and historical information. Discussion of reliability of estimates and usage demands points out that information in these areas has not been reported very thoroughly. Four definitions are given of construct validity and the models examined in relation to these definitions. All models are considered face or operationally valid, but comprehensiveness, discriminant validity, comparative validity and information validity need to be established for all models. A paradigm for predictive validity is put forward in which the criteria are early empirical results of fielded training and training devices. Discrepancy methods are suggested for relating predictive and prescriptive constructs based on SME estimates (sometimes bolstered by predecessor/similar training programs) to early empirical criterion standards. Followback and concurrent study designs may be effectively employed to capture predictive, comparative, and discriminant validity. Tests and performance measures, trade-off methodologies, and decision making under uncertainty methods are discussed. The chapter concludes with a discussion of generic vs. specific models and points out directions that need to be pursued before a generic model would be considered feasible. An integration of TDDSS and DEFT is suggested as the most promising lead for a generic model subject to further analytic and empirical testing.

Chapter 2

THE CONTEXT OF PREDICTIVE COST AND TRAINING EFFECTIVENESS ANALYSIS: TRAINING EFFECTIVENESS ANALYSIS (TEA) AND INSTRUCTIONAL SYSTEMS DEVELOPMENT (ISD) PROCESSES IN THE WEAPON SYSTEM ACQUISITION PROCESS (WSAP)

INTRODUCTION

Two general guidebooks have been issued by the Army for designing training programming and conducting cost effectiveness analysis of these training programs.

TRADOC provides guidance for conducting training effectiveness analysis in The TRADOC Training Effectiveness Analysis (TEA) System (May, 1979). It is a general guidebook for the types of TEA and the roles and responsibilities in these TEA. Instructional Systems Development (ISD) is a step-by-step conceptual framework for designing, developing and evaluating training programs. These two guidebooks chart the course for training development and CTEA within the military setting.

TRAINING EFFECTIVENESS ANALYSIS

TRADOC provides guidance for training effectivness assessment in training: The TRADOC Training Effectiveness Analysis (TEA) System (May 1979). This document prescribes responsibilities, policies and procedures which govern the operation of the TRADOC TEA System. The above mentioned guidebook is also available as TRADOC Training Effectiveness Analysis Handbook, TRASANA (1980). This handbook outlines procedures for conducting TEA. The step-by-step methodology is described to aid the analyst in what to do, and how to do it. Data sources are also suggested.

Table 2-1 shows the various types of TEA. Of particular interest to this project are the two methods which involve both cost analysis and TEA, that is the Cost and Training Effectiveness Analysis (CTEA) and the Training Development Study (TDS). These methods are both appropriate to predictive CTEA's during early phases of the Weapon System Acquisition Process (WSAP). Table 2-2 shows the roles and responsibilities for CTEA. Throughout this report we will use the term CTEA in a generic sense to incorporate TDS and distinguish among various phases of the WSAP.

Table 2-1

Types of TEA Described in TRADOC Regulation 350-4

Type	Description
1. Cost and Training Effectiveness Analysis (CTEA)	Conducted during the acquisition process in order to: Insure that Training Development (TD) processes (ISD Phases I, II, and III) are initiated early in the life cycle of hardware systems and are accomplished both in parallel and coordination with Combat Development (CD) processes during the acquisition cycle. Optimize soldier hardware subsystem interface. Insure that the appropriate level of scientific methods are used in the development of the training subsystem. Insure that all feasible training subsystem alternatives are considered. Optimize soldier training subsystem interface. Recommend the preferred training alternative for the preferred hardware system based on cost and training effectiveness. Provide decision-makers with more precise information at critical points in the acquisition process concerning the Total System comprised of the training, hardware, and other subsystems (TRADOC Pamphlet 71-8.)
Initial Screening Training Effectiveness Analysis (ISTEA)	Conducted after a system has been fielded to determine the relation between design effectiveness (ED) and actual effectiveness (EA) by analyzing soldier and teacher proficiency, attitude, and other training environment variables related to EA/ED relationship.
Training Subsystem Effectiveness (TSEA)	Conducted after a system has been fielded in order to determine if the performance gap is caused by or related to the various aspects of the training subsystem and the related subsystems, and identify potential solutions.
Training Developments Study (TDS)	Usually conducted after a system has been fielded but also is used preliminarily to the conduct of CTEA for developing system training devices and nonsystem training devices which are under separate Training Device Letter Requirement (TDLR). The TDS is designed to: Find the most cost-effective way to fix training subsystems found deficient during the conduct of a TSEA. Find the most cost-effective way to change training sybsystems which are not deficient but considered too costly or in need of revision.
Total System Evaluation (TSE)	Conducted on fielded systems when performance gap is partially due to personnel and logistical subsystems.

- Adapted from Knerr et al., 1984

Table 2-2

CTEA Roles and Responsibilities as Assigned by TRADOC

HQ TRADOC	Provides policy, direction, program review, and study approval.
Deputy Chief of Staff for Training (DCST)	Directs the TRADOC TEA System.
Deputy Chief of Staff for Combat Development	In coordination with the TRADOC DCST, insures OMA and RDT&E funds for the CTEA portion of the COEA are included in the programming of funds (DD Form 1498) for COEA and other related combat development studies (TRADOC Regulation 11-8). Functions as HQ TRADOC point of contact for TRASANA in CTEA matters as they pertain to the overall COEA effort. In support of the CTEA effort, provides a coordination link which facilitates TRANSANA's entry into the TSM, HQ DARCOM and PM loop.
U.S. Army Training Support Center (ATSC)	Serves as the TRADOC DCST point of contact (POC) for proponents in matters relating to CTEA study directives for developing systems and nonsystem training devices.
TRADOC Systems Analysis Agency (TRASANA)	Responsible for the TRADOC TEA Handbook which explains the how to procedure and methodologies for each type TEA. Conducts independent TEA efforts as directed by HQ TRADOC.
Proponent Service Schools	Serve as TEA study proponent.

Adapted from Knerr et al., 1984

ISD PROCESSES

The TRADOC pamphlet 350-30, Interservice Procedures for Instructional Systems Development, replaced Army-issued Regulation 350-100-1 Systems Engineering of Training which was based upon the early model, USAF manual 50-2, Instructional Systems Development, 1970. ISD prescribes design, development and evaluation of military instructional programs. ISD is a conceptual framework and sets the stage for models and training development design. It consists of five phases:

Analyze - A job analysis is conducted to identify tasks which require training. This job analysis is the basis for the rest of the ISD processes. The various aspects of task training are evaluated, such as task learning difficulty, probable consequences of inadequate performance, etc., including economic and time considerations. Job performance measures are identified. These determine job proficiency, evaluate training and maintain quality control. Existing courses are analyzed to determine if there are any which teach those tasks. After this analysis, tasks are assigned to an instructional setting depending on the capability and resources of the setting. The five instructional settings include: job performance aids, self teaching exportable packages, formal on-the-job training, installation support schools, and resident school. The capability and resources of instructional settings and the need for training are important determinants for selecting instructional settings.

Design - Learning objectives are defined for each task. The learning steps are outlined to master the learning objectives. Criterion-referenced tests are constructed for the various end of course tests and pre-tests. Sequence and structure are developed for learning objectives and steps and presented to the students.

Develop - The development phase of ISD classifies tasks into sub-categories. Decision matrices help determine which media and mode are to be used for lesson presentation. An evaluation of instructional materials is conducted.

Implement - Instruction is planned, implemented, and monitored to identify problem areas and/or changes in direction from the original plan.

Control - Internal and external evaluations are conducted to determine the efficiency of the instructional development effort and to evaluate weaknesses in the application of ISD. Revisions are suggested.

ISD does not analyze features within devices. It does not employ metrics for effectiveness or costs. It is not automated. The various steps in implementing ISD are not necessarily linear and may take place concurrently.

Since ISD is a conceptual framework, it does not provide a much needed formal model for training design, development or predictive CTEA. Knerr et al. (1984) note that ISD is a general guide and has not been followed systematically in the Army. Reliability and validity of predictive measures implied within the ISD framework have also not been addressed.

CTEA IN VARIOUS PHASES OF THE WSAP

Rosen et al. (1981) note that CTEA is sometimes used in a generic sense to include analyses conducted at all phases of the WSAP and sometimes in the specific sense of analyses conducted during the acquisition vs. the fielding phase of the WSAP in the Life Cycle System Management Model (LCSMM). Matlick et al.'s review identified the locations for four CTEA and updates in the WSAP (corresponding to the LCSMM) as follows:

LCSMM/WSAP Phase	CTEA
I Conceptual	CTEA I and update
II Demonstration and Validation	CTEA II and update
III Full Scale Development	CTEA III
IV Production and Deployment	CTEA IV (update)

The LCSMM events related to CTEA and the issues they must resolve are shown in Table 2-3. Rosen et al. (1981) note that these processes are relevant to training devices and training development in general, as each set of unique conditions allow, but are less clearly defined after a WS becomes operational.

The TRADOC Training Effectiveness Analysis Handbook TRASANA (1980) is a procedural guide that outlines a general methodology for conduct of CTEA's. The text describes the method step-by-step and suggests data sources.

The Training Acquisition Handbook (CORADCOM, 1980), a combined DARCOM and TRADOC set of information guidelines, reviews the ISD process and training acquisition in the LCSMM context. This document has given impetus to formal CTEA models by requiring that training formats other than written formats must be justified by a media analysis. Training device acquisition is given major attention in Chapter 6 of this document. Training devices and simulators (TD/S) are defined as panel displays, simulators, part-task trainers and full crew trainers. They are distinguished from training aids (descriptive charts, graphs, and audio/visual materials) and training equipment (operational equipment dedicated to training). Major devices and simulators are recommended for use when critical subject matter is too complex for verbal, symbolic or simple pictorial presentation or when it requires extensive hands on experience to develop requisite skills.

PM TRADE (1979) classifies training devices by their intended use:

1. Weapon system specific device
2. Non-system for use in one or more WS

More generic (non-system) TD/S, those at high development risk costing over $3 million, follow the general Army LCSMM. In this context, CTEA may be necessary only if a training system must be designed to manage the TD/S and/or a unique training system set up to teach, operate and maintain it. The Cost and Operational Effectiveness Analysis (COEA) serves as a CTEA.

Rosen's (1981) discussion brings out many other details of TD/S acquisition. Suffice it to say that a generic CTEA model must eventually deal effectively with development of training for a specific WS, TD/S, task level training, non-system training and institutional and unit training. This report reviews the existing CTEA models, their functions and areas of application and suggests directions for further model development.

Table 2-3

Cost & Training Effectiveness Analysis (CTEA) in the Life Cycle System Management Model (LCSMM)

LCSMM PHASE	CTEA	PRIMARY LCSMM EVENTS YIELDING DATA TO CTEA	ISSUES REQUIRING RESOLUTION
CONCEPTUAL	I	• MENS • INITIATION OF TRAINING PLANNING • INITIATION OF LOGISTICS SUPPORT PLANNING	• TRAINABILITY OF BASIC CONCEPT • COST OF TRAINING • TRAINING PROGRAM ELEMENTS TO BE INCLUDED OR STUDIED
	I A (update)	• LOA • ORGANIZATIONAL AND OPERATIONAL CONCEPTS	• TRAINABILITY OF ALTERNATIVE CONCEPTS • RELATIVE COST EFFECTIVENESS OF TRAINING PROGRAMS OF ALTERNATIVE CONCEPTS
DEMONSTRATION AND VALIDATION	II	• OAP • DT/OT I	• TRAINABILITY • NECESSARY REVISIONS OF TRAINING PROGRAMS • RELATIVE COST EFFECTIVENESS
	II A (update)	• PQQPRI • DCP(IPS)	• TRAINABILITY • TDR • RELATIVE COST EFFECTIVENESS OF REVISED TRAINING PROGRAMS
FULL-SCALE ENGINEERING DEVELOPMENT	III	• NET PLAN • AP • DT/OT II	• PERFORMANCE VERSUS STANDARDS • PERFORMANCE VERSUS HARDWARE AND TRAINING PROGRAM DESIGN, PERSONNEL SELECTION, ETC.
PRODUCTION AND DEPLOYMENT	IV (update)	• TRAINING PLAN UPDATE • DRAFT TRAINING PROGRAM • AP • DT/OT III	• COST EFFECTIVESS OF REVISED TRAINING PROGRAM

Source: Adapted from Matlick et al., 1980, p. II-41. See for details of these events.

Chapter 3

COST AND TRAINING EFFECTIVENESS IN ARMY MATERIEL ACQUISITION

INTRODUCTION

A number of concepts and models have evolved that relate WS acquisition to manpower, personnel and training. These systems frequently incorporate models specific to training. This chapter includes the following:

1. Trade-Off Resolution Support for System Acquisition (TORSSA)
2. Knerr's (1984) review of applicable models
3. HARDMAN
4. Coordinated Human Resources Technology (CHRT)
5. Acquisition of Supportable Systems Evaluation Technology (ASSET)

TRADE-OFF RESOLUTION SUPPORT FOR SYSTEM ACQUISITION (TORSSA)

Dawdy and Hawley (1984) outline a conceptual approach for the Army for early consideration of human performance (HP) requirements in conjunction with hardware and software design in the conceptual phase of the WSAP. They note that increasingly complex WS have resulted in (1) "skill creep," that is, higher demands placed upon the personnel who must operate, maintain or support them, or (2) WS that do not perform as intended or that cost more than projected.

These problems may be obviated at least in part by early consideration of HP requirements as part of WSAP so that the WS does not make unreasonable HP demands. HP subsystems requiring consideration include human factors, manpower, personnel and training (HMPT). All HP "subsystems" must be treated in interaction with each other and the hardware and software design.

The authors point out some of the inadequacies in HP technologies and recommend approaches that focus on an integrated, proactive, decision analytic framework that will enable trade-offs to be made between hardware/software designs and HP designs.

Integration among the HMPT factors is equally important. The approach must also have a "designer focus," that is "... the methodology must be tailored for use by personnel who actually are responsible for early system design decisions... and to provide appropriate design, such as training device specifications, document training plans, logistic support plans to users such as the Army Materiel Command and TRADOC."

<u>**The TORSSA Process**</u> is carried out in five phases, each with a number of stages:

1. Mission Resources Management Analysis (MRMA)
   - 1.1 Define threat situation
   - 1.2 Generate collective threat scenario
   - 1.3 Develop statement of required capability
   - 1.4 Identify resource requirements
   - 1.5 Specify resource requirements
   - 1.6 Develop employment prescriptions
2. Functional System Definition (FSD)
   - 2.1 Characterize system, subsystem and equipment components
   - 2.2 Specify operational concept

2.3 Allocate functions to hardware, software or human elements

2.4 Evaluate and refine function allocation elements.

2.5 Develop H/M interface specifications

2.6 Evaluate alternative system concepts

3. Manpower and Personnel Requirements Analysis (MRPA)

3.1 Specify system functional relationships

3.2 Define position requirements

3.3 Identify manning requirements

4. Training Cost-Benefits Analysis

4.1 Development of training program alternatives

4.2 Selection of a preferred training program

5. Trade-Off Resolution

5.1 Select preferred system in WS concept development phase

5.2 In operational testing phase, identify and eliminate performance shortfalls

A review of the TORSSA process outline suggests that it may obviate one of the recurring problems of HMPT in the WSAP, i.e., the "trailing" of HP designs behind hardware/software designs, where only "retrofit" solutions are possible, often at a high cost.

Dawdy and Hawley (1984, pp.22, 23) point out that:

"...At the early design stage, the authors' experience has indicated that it is sufficient simply to evaluate each action, decision, and transaction's HP demands along physiological, sensory/perceptive, and cognitive dimensions. The primary concern at this step is that a proposed design concept not exceed human physical, sensory, or information processing capacities. Standard human design guidelines can be used to determine the suitability of function allocation concepts with regard to sensory/perceptive and physiological requirements. Methods for assessing a design concept in terms of its cognitive demands are not as well developed, but several tentative procedures have been proposed (see Card, Moran & Newell, 1983; Wickens, 1984).

The objective of the allocation review step is to identify those design concepts likely to impose unacceptable HP demands. Alternatives that are judged to impose unreasonable HP demands are either eliminated from further consideration or refined to reduce their HP requirements."

Phases three and four, Manpower and Personnel Requirements Analysis and Training Cost-Benefits Analysis, follow jointly upon Phase 2. Trade-offs are considered in Phase 5, step 1 to select a preferred system.

The training cost-benefits analysis uses the Training Developers Decision Aid (TDDA)/Training Development Support System (TDDSS) model (Frederickson, Hawley & Whitmore, 1983). As noted elsewhere in this report, this model includes training program design and prediction in terms of expected performance and time.

The authors note that TORSSA has evolved over a period of years representing the integration of a number of research and development efforts and applications primarily concerned with HP considerations in emerging systems. Part of the methods were applied to the operator-software interface assessment for the Phase III Product Improved Hawk air defense missile system (Dawdy, Fullam, & Hawley, in press). Training cost-benefits analysis evolved from early-on Cost and Training Effectiveness Analyses (CTEAs) (Dawdy, Chapman, & Frederickson, 1981; Dawdy, Chapman, & Brett, 1981) and TORSSA's training design procedures parallel those used in the TDDA (Frederickson, Hawley, & Whitmore, 1983).

The authors also point out that TORSSA is consistent with DOD and U.S. Army regulations and guidance for WSAP, including basic systems acquisition procedures for the Defense Department prescribed in DOD Directive 5000.1, Major Systems Acquisition, and in DOD Instruction 5000.2, Major System Acquisition Procedures interpreted for the U.S. Army in AR 1000-1, Basic Policies for System Acquisition, and in DA Pamphlet 11-25, Life Cycle Systems Management Model for Army Systems, supported in turn by numerous clarifying regulations and pamphlets that prescribe in detail the requirements for treating specific aspects of the WSAP.

KNERR'S REVIEW OF APPLICABLE MODELS

This report was particularly concerned with training devices. Knerr et al. (1984a) point out that: "Cost and training effectiveness analysis provides information to decision makers on the characteristics of the developing system"; several CTEA are needed for the Life Cycle Systems Management Model (LCSMM), and that CTEA provides cost and effectiveness assessment of the alternative training systems to the various stages of LCSMM.

They point out that six input data situations may exist for the analyst:

1. No task list and no training program
2. Task list but no training program
3. Training program but no alternatives and no effectiveness data
4. Training program with effectiveness data but no alternatives
5. Alternative training programs but no effectiveness data for alternatives

6. Training program alternatives and effectiveness data for all alternatives

At the time of their report no one CTEA model appeared adequate for providing information to LCSMM for reaching decisions. A number of models reviewed suggested that each model may have some importance or can be applied at the different input levels to reach a decision. The models they reviewed were:

1. Training Efficiency Estimation Model (TEEM)

2. Training Consonance Analysis (TCA)

3. Training Effectiveness, Cost Effectiveness Prediction (TECEP)

4. Methods for the Analysis of Training Devices and Simulators (TRAINVICE)

5. Army CTEA methods in current use:

   a. DIVAD Gun CTEA
   b. Improved Hawk (Hawk PIP) Training Development
   c. Roland Training Development
   d. Improved TWO Vehicle (ITV) CTEA, and
   e. Diagnostic Rifle Marksmanship Simulators (DRIMS) CTEA

Since their report, additional models have been located and are reviewed in this report.

To aid the analyst in applying an appropriate CTEA to meet the demands of the data situation in the LCSMM, the models were grouped into six categories. (See Figure 3-1 .)

The CTEA model for Army WSAP was applied to the Position Location Reference System (PLARS) equipment but its validity has not been established. This CTEA model is a synthesis of a number of different prescriptive and predictive CTEA models.

A GENERATION OF TASK LIST	B PRESCRIPTION OF TRAINING PROGRAMS	C ESTIMATION OF EFFECTIVENESS	D COSTING OF TRAINING PROGRAMS	E COMPARISON OF TRAINING PROGRAM ALTERNATIVES	F RESOLUTION OF ISSUES
DIVAD GUN CTEA* HAWK PIP ROLAND	TEEM TECEP DIVAD GUN CTEA ATM	TEEM DIVAD GUN CTEA ATM TRAINVICE TCA	MATLICK ET AL. COST MODEL	TEEM TCA DIVAD GUN CTEA UKINS CTEA ATMOS TECEP	ITV CTEA UKINS CTEA TRAINABILITY ANALYSIS

Figure 3-1 CTEA Processes for Six Data Situations**

* Abbreviations are defined in the text.

**Adapted from Matlick, Berger, Knerr, and Chiorini, 1980.

HARDMAN

Although developed primarily for the Navy, this methodology has been applied to Army training and human resources problems within the last four or five years. Four main objectives of HARDMAN (Military Manpower and Hardware procurement) are:

1. Institute procedures to address manpower, personnel, and training requirements.

2. Provide the means for compliance with policy and acquisition procedures.

3. Develop tools and methods to assist program managers in considering the impact of system design on manpower, personnel and training.

4. Provide an assessment of manpower, personnel and training supportability before design decisions and resource allocations are made.

HARDMAN aids in determining human resource requirements. It identifies high resource drivers, operational and support concepts, and policies that generate human resource demands; and it provides information for determining human resource/equipment design trade-offs during the Weapon System Acquisition Process (WSAP).

HARDMAN analyses employ six steps which can be used to make global or fine distinctions between existing and proposed systems. These six steps are:

1. Establish Consolidated Data Base (CDB) - CDB is established by determining the CDB requirements, identifying and selecting relevant data sources, establishing CDB structure and format, performing systems analysis including determining the reference system and baseline system, and establishing an audit trail of the analyses.

   The reference system is a reconfiguration of components and existing equipment systems. The baseline system consists of this reconfiguration with new low risk technological innovations. The CDB data are extrapolated from the reference systems and WSAP. The accuracy of data depends on the assumptions made about the reference systems.

2. Determining Manpower Requirements - This step involves the user inputting data concerning the maintenance concept for the reference system, operational requirements, repair concept, etc.

3. Determining Training Resource Requirements - Training Resource Requirements Analysis (TRRA) is used in the estimation of costs and resources for the predecessor, reference, and baseline systems with certain limitations such as no estimation for civilian and officer training. Three major steps are carried out to establish the training portion of the CDB, document training programs for the predecessor, reference, and baseline systems, and determine additional training requirements. Training programs that are most closely related to the reference and predecessor systems are identified for new and modified components. Tasks are categorized in the reference and other systems according to their action verbs. These are analyzed using the ISD and Training Effectiveness Cost Effectiveness Prediction (TECEP) methods. TRRA then identifies general media types. TRRA is being modified to incorporate a more detailed level of analysis.

4. Determine Personnel Requirements - Analyses in this step involve establishing the personnel portion of CDB personnel pipeline flow characteristics, and final personnel requirements.

5. Conduct Impact Analysis - This step calculates the manpower, personnel, and training requirements of the proposed system, and compares them to the projected supply. New requirements for skills, training and resources and other human resources demands are highlighted. This step has not been validated.

6. Perform Trade-off Analysis - This step involves the consolidation of requirements and resources, determination of solutions and initiating procedures to analyze proposed solutions.

The method has not been empirically validated but has been applied to several Navy and Army Systems. It is the procedure most widely used by the Navy for MPT estimation.

COORDINATED HUMAN RESOURCES TECHNOLOGY (CHRT)

According to Knerr et al. (1984), the combination of the following five techniques for estimation of MPT form CHRT:

1. Instructional System Development (ISD)
2. Job Guide Development (JGD)
3. Maintenance Manpower Modeling (MMM)
4. System Ownership Costing (SOC)
5. Human Resources in Design Trade-offs (HRDT)

A more detailed description of CHRT can be found in a report by Goclowski, King, Ronco and Askren (1978a, 1978b, 1978c). This model was developed for the Air Force and has been refined into ASSET.

ACQUISITION OF SUPPORTABLE SYSTEMS EVALUATION TECHNOLOGY (ASSET)

Knerr notes that ASSET can be applied during all phases of the acquisition process to:

1. Provide assessments of costs, human resources and logistics resources required for support of the weapon systems.
2. Coordinate the development of the training programs and the technical manuals used in these programs.
3. Give consideration to the support systems and human resources and their impact on the design of the weapon system.

The three components of ASSET are:

1. Consolidated Data Base (CBD). This contains all the information required to analyze the human resource and support impact during the acquisition process of the weapon system. The data bank is continually updated.
2. Eight analysis procedures.

3. Eight analytical computer models.

SUMMARY

The weapon system acquisition process includes considerations beyond training, particularly tradeoffs among hardware, manpower, personnel and training. The TORSSA concept articulates the problems and an approach for the Army for further development. An Army handbook, Sources of Information on Integrated Personnel and Training Support Planning: A Handbook for TRADOC Systems Managers (STEPS), describes the major events of the LCSMM including the training plan requirements.

Knerr et al.'s (1984a) review pointed out data input requirements for various models and reviewed the existing CTEA models and applications emphasizing CTEA for training devices.

HARDMAN was developed for the Navy but is being applied to Army MPT analyses. The remaining models were developed for the Air Force and as far as we know have not been adapted to the Army.

Chapter 4

PREDICTIVE COST AND TRAINING EFFECTIVENESS ANALYSIS (CTEA) MODELS

INTRODUCTION

Predictive models attempt, as the term is used in this report, to predict the performance of trainees, training time, instructor time, and, in some cases, training resource requirements. In contrast, prescriptive models are concerned with combining media and method characteristics into a "figure of merit" for a training program design. In one case, TDDSS/TDDA, the models are complementary, serving both predictive and prescriptive purposes.

Thus far, five approaches to predictive CTEA models have been generated for conducting Cost and Training Effectiveness Analysis (CTEA) for Military training programs. These predictive CTEA approaches are:

1. Training Development Decision Support System (TDDSS)

2. HARDMAN Training Resources Requirements Analysis (TRRA)

3. Job Skills Education Program (JSEP) CTEA for Computer Based Instruction

4. Comparison based prediction of cost and effectiveness of training devices

5. Training Device Effectiveness (TRAINVICE) and Device Effectiveness Forecasting Technique (DEFT) models

TDDSS addresses entire training programs but has also been used to provide input to training devices and simulator development.

Kraft and Farr's (1984) JSEP is using methods of decision making under uncertainty applied to the distribution of Computer Based Instruction (CBI) systems. The decision making under uncertainty methods have been used widely in weapon systems acquisition, manpower and personnel, but not in training prediction or prescription. Comparison based prediction has been applied to training device effectiveness and costs.

TRAINVICE and DEFT address the transfer effectiveness of training devices and simulators. No corresponding cost models have been developed.

TRAINING DEVELOPMENT DECISION SUPPORT SYSTEM (TDDSS)

This methodology was developed by Hawley & Fredricksen (1983) and reviewed recently by Knerr et al. (1984). TDDSS is a continually evolving methodology. It is an automated model for training design and evaluation which includes cost benefit analysis. TDDSS contains a data base, a model base, and a software system that includes data base and model base management. It is based on ISD methods. TDDSS does not have rigidly defined sequential steps that must be followed to make use of the system.

The training design portion of TDDSS (TDDA) is concerned with "what to train," "where to train" and "how to train" (Hawley & Fredricksen 1983, p. 398).

The input information for the task and job analysis is obtained from Expert Job Performers (EJPs). EJPs have typically had experience in operational job assignments in the primary job position. They have had hands-on experience with the kinds of tasks and jobs typically specified for training programs. The training design portion of TDDSS has

3 phases:

1. Function Analysis - to generate a task list and to subject the approved tasks to a criticality analysis along ten criticality dimensions, such as learning difficulty and frequency of performance.

2. Task Analysis - a task hierarchy is generated with skills specified, so the focus is on training.

3. Learning Requirements Analysis - identifies the special features of the task, functional learning requirements, training constraints, and resources. The functional learning requirements produce information on the media and devices appropriate for the task and how to prepare the training content. The analysis emphasis is on the characteristics of the learners.

The Training Evaluation portion of the TDDSS utilizes a general Multiattribute Utility Measurement (MAUM) routine. The MAUM routine is described by Dawdy and Hawley (1983) as follows:

1. The worth of each task selected for training is determined. A list of task criticality factors is selected to define the bases of training worth. Each of the criticality factors selected is assigned a relative importance weight (Rj) which ranges from zero to one. Next, responses to each criticality factor are assigned utility ratings which range from zero to one. The resulting utility values for each task (Uij) are multiplied by their respective importance weights to obtain task training worth indices (Wi):

$$Wi = \sum_{j=i}^{n} Uij\ Rj$$

2. The second step involves estimating effectiveness of each Training Program Alternative (TPA) in training each task. A group of training experts reach a concensus regarding the percentage of trainees that can be expected to reach criterion on each task with each TPA. Various

aspects of each TPA are considered to arrive at training effectiveness estimates (Eki) where K represents TPA and i represents the task to be trained. After making these estimates, the experts consider again those tasks for which the Eki are either under standard (typically, 90% of the trainees should meet the standard stated in the performance objective) or over standard.

In case of tasks where Eki are under standard, estimates have to be made of extra time required to raise trainee standards. For tasks in which Eki are over standard, estimates are made of how much time could be cut from the training plan and still meet the standards. These time increments and decrements are denoted t+ and t-.

Aggregate effectiveness estimates are computed by summing the product of task-level training effectiveness estimates and indices of training worth across tasks.

$$TPE_k = \sum_{i=1}^{n} W_i E_{ki}$$

TPE_K = estimated effectiveness of the the Kth TPA

Wi = training worth rating of ith task

Eki = estimated effectiveness of the kth TPA in training the ith task.

Wi is scaled to range between zero and one and sum to unity; Eki falls in the interval 0 to 100; and TPE also ranges in the interval 0 to 100.

3. t+ and t- estimates are reviewed to determine whether more or less training time is required for each TPA to meet training standards on each task. The total estimated "slack" time is obtained by:

$$TS_{ki} = \sum_{i=1}^{n} t_{ki} D_{ki}$$

t_{ki} = t+ or t- value for the Ki ith task on the kth TPA.

Dki = 1 if more training time is required or

-1 if less training time is required

The components of the Training Evaluation are as follows:

Resource Projection - This is a structured, overall analysis of the resources required for the Training Program Alternatives (TPAs)

Cost Estimation - This model computes a cost estimation for each resource projection output.

Benefits Analysis - A general (MAUM) routine is used in this analysis. This routine accepts four dimensions - Model defined training modules, model defined TPA's, user defined evaluation dimensions (primary dimensions on which the TPAs are evaluated), and user defined contextual dimensions (additional analytic evaluation for the TPAs). An index of merit is obtained for each TPA from the analyses. This indicates the overall benefit of each TPA.

Trainability Analysis - This analysis aids the user in assessing the ability of each TPA to cover all the requirements of the training modules.

Cost-Benefit Integration - This analysis identifies the "... best training program which is affordable" (Hawley & Fredricksen, 1983 ,p. 401).

Alternative Selection - This is a review step to ensure that the preferred TPA is indeed the best selected.

The final TPA selected is the one which addresses all the important training issues and is within certain cost limits.

HARDMAN TRAINING RESOURCES REQUIREMENTS ANALYSIS (TRRA)

This model was reviewed in Chapter 3 in relation to developing Weapon Systems WS. It can in some respects be considered a predictive model.

JOB SKILLS EDUCATION PROGRAM: PREDICTIVE COST AND TRAINING EFFECTIVENESS ANALYSIS

Kraft and Farr's (1984) CTEA for the Job Skill Education Program (JSEP) for Computer Based Instruction (CBI) is in its developmental stages. The authors have outlined a guide for conceptualizing predictive CTEA specifically for JSEP CBI, as they note that the nature of these programs rendered other kinds of CTEAs inappropriate.

The analysis is to obtain the optimum mix of site locations, hardware, software and personnel to produce the highest achievement in a given period of time at selected Army posts in the continental United States. This model is eventually to be tested at Fort Rucker, Alabama. Further refinements of the model will be made after the data analysis has been completed.

The authors define CTEA as "... a methodology that involves documentation of the comparative effectiveness and costs of alternative training systems for attaining specified performance objectives" (Kraft & Farr, 1984, p.1). Thus the application of CTEA involves mutual consideration of performance variables, the values of which are subjectively determined,and related costs.

The report discusses the proposed JSEP-CTEA procedures as follows:

The Classification System of JSEP

The classification system of JSEP-CTEA is used to (1) define operational and functional constraints,(2) define the training configurations, and (3) to qualify the advantages and disadvantages of each design alternative for site-specific requirements. These efforts are to be accomplished by estimating training effectiveness of various training configurations at various site scenarios.

Cost Analysis

The basic input-output model for linking costs to system output criteria in JSEP-CTEA is outlined as:

Input	Output
Research and Development	Cost/Terminal Hour
Investment Costs	Cost/Instructional Hour
Operative Costs	Class Hours/per Training Center
Life-Cycle Costs	Hourly Cost/Graduate
	Direct Cost/Graduate, etc.

These input-output variables will be combined with data on:

1. Number of students/site
2. Student flow rate
3. System usage ratio (high/low)
4. Rate of completion of assigned lessons (deficiency ratio)
5. Average number of hours of instruction per student to completion

The proposed JSEP-CTEA Costing Process will start with a version of the basic cost model and will then carry out the following steps:

1. Develop cost methodology
2. Determine cost data requirements and identify cost data sources
3. Obtain cost data
4. Perform cost data analysis, including sensitivity studies
5. Determine the cost of each alternative
6. Describe non-quantifiable elements of training cost
7. Integrate CTEA with a decision oriented multi-criteria utility analysis

All of the above will depend upon the availability of data. Also to be investigated is whether benefits which cannot be expressed in dollars can be analyzed using rating scales and evaluated in terms of "equivalent dollars." Sensitivity analyses will be performed for a number of variables.

Aspects of Computer Based Instruction (CBI) will be investigated at the U.S. Army Forces Command (FORSCOM) and U.S. Army Training and Doctrine Command (TRADOC) sites. These aspects include total training hours per student and instructor student ratio.

Cost-Effectiveness Indicators are:

1. Cost/instructional hour = $\dfrac{\text{sum of direct cost/program costs}}{\text{total class hours available to the student}}$

2. Hourly cost/graduate = $\dfrac{\text{cost per instructional hour}}{\text{no. of graduate students per unit of time}}$

3. Direct cost per graduate = hourly cost per graduate x hours of instruction

Estimating future costs will be done by:

1. Sensitivity analysis using a high, medium and low for total system cost for each of the three alternatives.

2. A contingency analysis to show how a change in the environment impacts upon the ranking of alternatives or major change in criteria.

The technique of discounting or present value will also be incorporated in the cost analysis.

Site populations at FORSCOM and TRADOC will be the basis for site requirements analysis. Six scenarios will be drawn up based on:

1. Assumed data on throughput levels

2. Number of terminals installed

3. Number of instructors per site

4. Number of hours of instruction

The size of each system is expected to relate directly to the system's usage efficiency. The highest efficiency rate is obtained when maximum number of CBI terminals are installed and are operated 100% of the scheduled usage time. Since this is not the case in the real world, efficiency is expected to depend upon initial development cost, probable course usage and local utility rates.

Estimation of Training Effectiveness

Eight steps involved in Training Effectiveness Analysis (TEA) are outlined which would lead to a data base for comparing alternative systems:

"1 Measures of Training Effectiveness (MOTE) are defined and identified, and are used to answer training-performance related essential elements of analysis (EEA). Measures of training resource requirements (MTRR) are also identified, defined and selected to answer resource related EEA.

2. Data elements required to formulate MOTE and MTRR are identified.

3. Requirements of TEA studies are defined and identified.

4. Expected sources of data to answer MOTE and MTRR are identified.

5. Appropriate methodology to conduct studies and tests are developed.

6. Tests and studies are planned and data are collected to support CTEA.

7. Analysis of data is performed to determine if training effectiveness MOTE and MTRR have been answered.

8. Sensitivity studies are conducted using the data and results of data analysis."

JSEP-CTEA will be supplemented with Arrayed Criteria

Analysis (ACA). The steps in ACA are:

1. Construction of decision relevant goals and identifying criteria.

2. A preselection of alternatives which can qualify for consideration.

3. Description of how these alternatives affect individual criteria. This gives the data for conversion of outcomes into unweighted effectiveness scores.

4. Weighted coefficients will be established.

5. Integrated index by way of using effectiveness index with base data for each alternative."

The primary, intermediate and the overall targets are developed for each course of action. A target matrix is developed and incorporates all the alternatives. This forms the basis for weighting the various elements in the target levels, developing coefficients that reflect values placed on target level elements by major commanders and for deriving a fused utility index from these coefficients.

COMPARISON-BASED PREDICTION OF COST AND EFFECTIVENESS OF TRAINING DEVICES

Klein Associates' (1985) comparison-based prediction of cost and effectiveness of training devices (TD/S), according to the authors, is a conceptual approach which can be applied to TD/S early in the design sequence. This method does not require operational data from the system under design and can operate with information from other sources similar to the system. CBP utilizes structured expert opinion and data that are available from similar devices. 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..." (Klein 1985, pp. 1-4).

Thus, this method is a process whereby an inference of the effectiveness of a device is made from the effectiveness of another similar device while making adjustments for the key differences. This use of analogy allows the evaluators to concentrate on only the most important variables. The methodology is described as follows:

Elements of the CBP Methodology

1. Target Case A
2. Target Variable: T
3. Target Value: T (A)
4. Subject Matter Expert (SME)
5. Comparison Case(s): B
6. Causal Factors (from which high drivers are selected)
7. Scenario
8. Strategy
9. Comparison Value: T(B)
10. Audit Trail

Steps in Using CBP

Phase I: Set up the problem

1. Specify the device (A) for which cost effectiveness is being predicted.

2. Define the measure (T) of that cost or effectiveness. This is the variable to be predicted.

3. Identify the major causal factors (high drivers) that affect T(A).

4. Define the context for prediction.

Phase II: Select specific resources

5. Identify comparison devices.

6. Examine the CBP strategies to select the most relevant one.

7. Choose knowledgeable subject matter experts.

Phase III: Collect the data

8. Determine with the SME the comparison value T(B) which may already be known.

9. Examine the differences between A and B, and estimate the effect of these differences on T(B).

10. Adjust the value of T(B) to account for the differences between A and B.

Phase IV: Make the prediction

11. Determine the value for T(A) from this adjustment.

12. Document the process to leave an audit trail. This aids in evaluating this decision or revision as the development takes place.

The steps outlined above for using CBP are not to be taken as rigidly sequential.

Setting up the Problem

Klein indicates that it is easy to specify variables in cost prediction, but not in predicting effectiveness of training. Training effectiveness needs to be specifically defined, and specific variables need to be identified. A checklist of items to consider in selecting predictor variable T(A) is as follows:

1. What do you need to predict?
2. Are there standard measures?
3. Does T(A) need more than one measure?
4. Who will use this result, and how?
5. Does the measure reflect training device use?

6. How will T(B) data be obtained?

General Measures for Relative Comparisons

Measures of Cost	Measures of Training Effectiveness
investment costs	accuracy
operations & support costs	recall
instructor	speed of performance
facilities	transfer of training
maintenance	savings
life cycle costs	recognition
	performance on secondary tasks
	effort/efficiency
	number of wins (gaming tasks)
	number of instructors needed
	amount of supplementary actual equipment training
	skill decay curve
	time to criterion

The selection of T can be influenced by the availability of T(B) for comparison. The list of possible comparison cases (B's) does not have to be shortened until a final decision is made with an SME. At times, a whole scenario of events may have to be created within which comparison data may have to be estimated for a detailed measure.

Listing the Causal Factors (High Drivers)

The five to seven major factors which account for the majority of differences in value from one scenario to another can be termed causal factors or high drivers. It is important for the SME to assess the impact of differences of these factors between the actual case and the comparison case on the variable T. Other minor causal factors are not important because their impact is minimal. The SMEs identify these causal variables and note how they differ in value in the scenario from the actual device. Also, differences in the context, such as the level of student to be trained, will have sizable effects. Features such as student motivation, feedback mechanism, instructional quality, etc. will have significant impact on T.

Some Causal Factors Affecting (T) of Training Device Effectiveness

Physical fidelity
visual, audio, feel of controls

Device utilization
general ease of use, set up ease, equipment characteristics, frequency, computer power

Training value
instructional aids, feedback, familiarity, relaxation, motivation

Functional fidelity
gunner training, procedures-automatic, range of tasks, variety and quality of training problems, lead training, malfunctions, trainee level, task simplicity

Task characteristics
task simplicity, crew coordination

Trainee level

Miscellaneous
cost, range

Constructing the Scenario

Predictions are always made in some context. The conditions under which the device will be used needs to be known. Thus, the scenario has been incorporated as a formal element in the CBP method. The scenario must, at least, include the high drivers selected for this situation. Also, it must provide a specification of the conditions under which the device will be used.

Some factors to consider in developing the scenario are:

1. Trainees - who they are, their experience level, and whether they are similar/dissimilar to the current trainees.

2. Task - what is to be trained, whether the task is generic or specific and what the criteria are for learning.

3. Program - what equipment is used and where it is administered. Options in the unit, if the course is in a school and how it will be modified with the use of a training device.

Choosing the Comparison Case

A list of comparison cases needs to be constructed, and information gathered on them. An SME might be needed as a consultant for this task. A good comparison case would have a high degree of similarity on the high drivers. Physical fidelity and task type are considered two high drivers which should, ideally, be matched very closely. The use of an SME is very important in identifying the best possible comparison case.

Selecting the Expert

SME's are selected on the basis of 1) familiarity with the case, 2) high degree of familiarity with respect to high drivers and 3) readily attainable data for the comparison case. The selected SME must be able to conceptualize T(B) as the analyst has defined it. The SME's must have credibility. Several SME's can be used at different points in the CBP method, according to the need for expertise in the given area.

Alternative Strategies of Application of CBP

CBP application can be structured in several ways depending upon time constraints, number of comparison cases, availability of data, and identification of SME's.

1. Global strategy: One SME is interviewed and presented with all relevant data on A including a list of high drivers. The SME makes a prediction for T(A) based on his/her knowledge of T(B). This may be an actual value or a general statement.

2. High driver strategy: The SME details how A and B differ from one another. With a checklist of high drivers, the SME compares the two devices on these high drivers and how much difference they effect. The sum of these estimates is then calculated.

3. Multiple comparison strategy: Several comparison cases are initially used, then the choice is narrowed down to two or three.

4. Convergence strategy: Use of multiple comparison strategy as well as use of SME's multiple strategy. When using multiple comparisons, the SME's should be asked to rate only the device with which they are familiar. If they are experienced with more than one, the list of causal factors should be reduced to make it less confusing.

Table 4-1

Advantages and Disadvantages of CBP Strategy Options

Strategy	Advantages	Disadvantages
Global	low resource demands	weak prediction less explicit audit trail
High driver	explicit audit trail causal factors evident	difficult to use with several SMEs and/or multiple comparisons
Multiple Comparison Cases	structured predictions	requires more time and availability of multiple comparison cases
Multiple SMEs	broader input cross-checks possible	requires more time and SMEs. Resultant prediction is complex

5. Cumulative strategy: The SMEs can be added and interviewed one by one until enough agreement is achieved.

Collection and Analysis of Data

The interview can be conducted face-to-face or over the telephone. The following information should be available for the interview:

1. A description of A, the training device
2. A clear statement of the prediction target T(A)
3. A scenario for T(A)
4. A list of potential comparison cases, B's
5. The checklist of high drivers
6. Baseline data for T(B), if it isn't to be estimated by the SME
7. A glossary of CBP terms and definitions

A data form should be kept on the SME's background relevant to this topic.

Documenting the process

A brief description of SME, a description of the comparison cases, the ways in which comparison cases differ from the target case, the magnitude of these differences and the value obtained for T(B) should be documented. This documentation is needed for reviewing the procedures and improving later predictions. The CBP process can be applied to cost predictions as well, using the same principles.

CBP Experience and Validity

According to Klein (1985) CBP has several characteristics which make it useful to apply in the early stages of training device development. It does not require extensive data from the device about which predictions are to be made; predictions are derived from operational experience; it uses structured expert judqement; it asks for judgements relative to similar cases; and it leaves an audit trail of the prediction process.

These advantages, however, are countered by some drawbacks. CBP requires data from specific related cases and prediction is more accurate if a formal, reliable model is used. The comparable cases must be known to be successful, thus a limitation is imposed on the number of similar cases from which data can be utilized. CBP requires expert judgment but it may be equivocal as to who the expert is in the field.

According to the authors, CBP has been developmentally tested in predicting such measures as time saved in training and effectiveness of training. CBP has been applied to automotive maintenance trainers, VideoDisc gunnery simulators for tanks (VIGS), and trainers for self propelled howitzer operations and maintenance (HIP). CBP methodology has been compared to actual test results of effectiveness of training devices at George Mason University. The results, as yet not published, yielded a correlation of .90 between CBP predictions and test results (Klein, 1985). Another study noted that training personnel showed greater confidence in predictions using the CBP methodology as compared with their own unstructured judgments.

TRAINING DEVICE EFFECTIVENESS MODELS AND THE DEVICE EFFECTIVENESS FORECASTING TECHNIQUE (DEFT)

The Training Device Effectiveness Model (TRAINVICE) was originally conceptualized by Wheaton, Fingerman, Rose, and Leonard (1976). TRAINVICE models attempt to predict training device (TD/S) transfer to performance in operational settings. Three other TRAINVICE models were generated by Hirshfield and Kochevar (1979), Narva (1979 a,b), and Swezey and Evans (1980). In 1984, the Device Effectiveness Forecasting Technique (DEFT), a reconceptualization of TRAINVICE, was developed (Rose and Wheaton, 1984). Further details on the background of TRAINVICE models can be obtained from recent reviews by Knerr, Nadler and Dowell (1984) and Tufano and Evans (1982). The TRAINVICE models use a structured interview using SMEs to evaluate a training-device based training system. DEFT is contained on a menu-driven computer program which aids the analyst in rating a training device.

Knerr et al. (1984) and Tufano and Evans (1982) outline the components of TRAINVICE as follows (see Table 4-2):

1. Coverage requirements: Determine which skills and knowledge warrant training, i.e., those in the operational setting which should be trained in the device (TRAINVICE C and D).

2. Task communality: Assess the overlap between subtasks in the operational setting and in the training device (TRAINVICE A); Determine whether the subtask is trained in the device (TRAINVICE B); determine whether skills and knowledge in the operational setting are trained in the device (TRAINVICE C and D, called coverage analysis).

3. Physical similarity: Judge how well displays and controls are represented in the training device (TRAINVICE A and B). If the information is represented in the device, it is rated on the degree of similarity with the operational equipment.

4. Functional similarity: Determine the information processes of the operator (type, amount, and direction of information) for each display and compare the information requirements in the device to those in the operational equipment (TRAINVICE A and B). If the information is represented in the device, it is rated on the degree of similiarity with the operational equipment.

5. Learning: Rate learning deficit, the difference between the trainees' repertory of skills and knowledge and the level required to perform a subtask (TRAINVICE A); estimate how much the trainees have to learn, weighted by the time required to train them (TRAINVICE B); assess the criticality and the difficulty of each skill (TRAINVICE C); assess the degree of proficiency required and the learning difficulty of each skill (TRAINVICE D).

6. Training techniques: Use TECEP categories to determine task categories and rate how well the device implements the associated TECEP learning guidelines, then average the lowest of the ratings in each task category (TRAINVICE A); assess physical and functional characteristics as to how well they implement the learning guidelines from TECEP, as expressed in ISD (TRAINVICE C and D).

DEFT reorganizes the TRAINVICE components and literature into different rating scales with greater face validity. Three levels of analyses are available ranging from global to detailed: DEFT I (global), II (task level), and III (subtask level). These levels of analysis may be used at various phases of WSAP and training device development, depending on the information available. Four major analyses have to be conducted at each level:

1. Training Problem - (TP) is an estimate of the magnitude and difficulty in overcoming the performance deficit. Level and type of proficiency associated with the training objective and trainees' level of knowledge prior to using the device. This is analogous to the "learning" component of TRAINVICE.

2. Acquisition Efficiency - (AE) takes into

Table 4-2
TRAINVICE and DEFT Components and Variables*

CLASS OF VARIBLES	A Wheaton et al.	B Honeywell	C Narva	D Swezey and Evans	DEFT
Coverage Requirement	----	----	Coverage Requirement Analysis	Coverage Requirement Analysis	----
Communality Transfer Efficiency (DEFT)	Task Communality	Task Commu-nality	Coverage Analysis	Coverage Analysis	Analysis of Transfer Principles
Physical/ Functional Similarity	Physical Similarity Analysis	Physical Simi-larity index	------	------	Physical Similarity
Transfer Problem Analysis (DEFT)	Functional Similarity Analysis	Func-tional Simila-rity Index			Functional Similarity Analysis
	Similarity Score				
Learning TPA. (DEFT)	Learning Deficit Analysis	Skill & knowledge Require-ments	Training Criticality Analysis	Proficiency Analysis	Performance Deficit
Training Problem (DEFT)		Task training Diffi-culty Index	Training Difficulty Analysis	Learning Difficulty Analysis	Residual Deficit Learning Difficulty
Training technique	Training Analysis	----	Physical Characteris-tics Analysis	Physical Characteris-tics Analysis	Analysis of Training principles and Instruc-tional features
Acquisition Efficiency			Functional Characteris-tics	Functional Characteris-tics	

*Adapted from Tufano and Evans, 1984

account the quality of training provided by the device and the extra device variables which affect acquisition of skills required to meet training objectives. Assessment is made of training principles and instructional features of the device. This is similar to the training techniques component of TRAINVICE.

3. Transfer Problem Analysis - This is an estimate of the performance deficit that the trainees bring to the parent equipment after graduating from the training device. It assesses residual deficit, and difficulty in overcoming this deficit. Also, physical and functional similarity between the device and equipment is assessed. This corresponds with functional similarity, physical similarity and learning components of TRAINVICE.

4. Transfer Efficiency Analysis - This is concerned with measuring the transfer of skills/knowledge learned from the device to the equipment. The analysis is an evaluation of the transfer principles that the device incorporates. This is similar to the task commonality component of TRAINVICE.

The components of TRAINVICE and DEFT are summarized in Table 4-2. Each subtask or skill is rated for each TRAINVICE component. The scales for rating them vary among the different components and TRAINVICE models. The Transfer Potential Index is computed by using the following formula:

$$T = \frac{Ci \times Si \times Ti \times Di}{n \sum_{i=1} Di}$$

T = Transfer

C = Task

S = Similarity Index

T = Training Techniques Index

D = Learning Deficit Index

For DEFT, the analyst provides responses to rating scales for each of the DEFT components, the number and kind of which vary according to the DEFT Level. DEFT I requires 8, DEFT II requires 13 (same at task level) and DEFT III requires 35 ratings (same at subtask level). The ratings are entered into the computer according to a predetermined sequence.

For example, for Training Problem ratings on DEFT I, the analyst has to assess the proportion of skills/knowledge required to meet training objectives and the difficulty in acquiring them on a scale of 0 to 100. For DEFT II, the analyst has to rate each task stored in the data file on a scale of 0 to 100 for the question addressed in DEFT I. Most scales are rated from 0 to 100. Seven indexes of effectiveness are generated from the analyses and are shown in Table 4-3. A summary index is provided for the acquisition problem, training efficiency, transfer problem, and training efficiency analyses. The two training indexes and the two transfer indexes are then aggregated into acquisition and transfer indexes; then the two are combined to yield a single index of training device effectiveness.

The indexes for DEFT II and DEFT III are identical to DEFT I. The difference is that the ratings are averaged over the number of tasks/subtasks evaluated.

TRAINVICE and DEFT both employ a very detailed level of analysis for effectiveness prediction and have the advantage of organizing data and making quantifiable predictions. Both analyze components of subtasks/tasks and evaluate features within a device rather than the selection of methods and media in general. TRAINVICE has been automated

Table 4-3

DEFT I Indexes

Training Problem (TP) = $\dfrac{\text{Performance deficit (PD)} \times \text{learning difficulty (D)}}{100}$

Ranges from 0 to 100

Acquisition Efficiency (AE) = $\dfrac{\text{Rating}}{100}$

Ranges from .01 to 1.00

Acquisition (A) = $\dfrac{\text{Training Problem (TP)}}{\text{Acquisition efficiency (AE)}}$

Ranges from 0 to 10,000, with a low value indicating and "effective" device.

Transfer Problem (TRP) = $\dfrac{\text{RPD} \times \text{RLD} + \text{AD}}{100}$

Where

RPD = Residual Performance Deficit

RLD = Residual Learning Difficulty

AD = Additional Deficits or Physical Similarity, Functional Similarity

Ranges from 0 to 200

Transfer Efficiency (TT) = $\dfrac{\text{Rating}}{100}$

Ranges from .01 to 1.00

Transfer (T) = $\dfrac{\text{TRP}}{\text{TT}}$

Ranges from 0 to 20,000, with a low value indicating an effective device.

Total Effectiveness (Σ) = A + T

by the Army Research Institute and the Human Resources Research Organization (HumRRO) has produced workbooks for users ratings (Knerr et al., 1984b). Neither DEFT nor TRAINVICE, however, address cost analysis.

<u>Applications, Reliability, Validity and New Developments</u>

Knerr et al. (1984) reviewed various applications of TRAINVICE. They report that Wheaton, Rose, Fingerman, Leonard, and Boycan's (1976b) study using the Burst on Target (BOT) trainer and other devices, the prediction was consistent with the actual measures of performance. Harris et al. (1983) applied the TRAINVICE models to four tank gunnery devices. They found that the rater agreement was high on the coverage and communality ratings, moderate on the physical and functional similarity ratings, and low in the learning and training technique ratings.

Faust, Swezey and Unger (1984) evaluated the four TRAINVICE models for reliability, validity, and convenience of application in the field setting. The models were applied to two Army maintenance training and evaluation systems (AMTESS) simulators. The predictions of the models were compared to the actual transfer of training scores of the students. There were 63 students and nine analysts.

Results showed that (1) a summary index is misleading due to quirks in the mathematical computations, (2) the task-level and summary indexes were reliable ($r=.33$, $p < .05$), and (3) the Wheaton et al. model and the Swezey and Evans models are more predictive ($r=.34$, $p < .05$) at the task level, with the latter being more efficient. This study is a pilot study of reliability and validity as only a few tasks (34) were studied and only a few analysts (9) rated the models. All the models tended to be cumbersome and time consuming to apply. It is our understanding that further studies of this nature are underway.

Rose and Martin (1984) conducted an analytic assessment of DEFT. DEFT was evaluated in the following areas:

1. interpretation of output
2. sensitivity analysis
3. comparison of outputs
4. stability
5. interrater agreement

The first four analyses were subjects of Monte Carlo analysis. Eight input variables were used and uniform distributions were selected as no empirical distributions exist. Five thousand trials were selected for each analysis.

Results for "Interpretation of Output" show that DEFT showed low variances when normal distributions were used based on the authors' familiarity with the training devices. Sensitivity analysis showed that all scales are weighted equally and have equivalent effects on the total score, except efficiency scales which have a larger effect. Analysis of "Comparison of Outputs" showed that differences between various outputs under different conditions were distributed normally. Stability of the magnitude of +5 or +10 was found for all five DEFT model outputs.

Six raters were used to determine the degree of interrater agreement using DEFT. The raters evaluated 3 training devices: MK-60 gunnery trainer, burst-on-target trainer, and a maintenance procedures simulator.

Data showed that DEFT I and III are internally consistent. Of the 84 pairs (3 devices x 4 raters x 7 indexes) of DEFT I & III indexes, 70 (83.3%) were within 20 points of each other and half were within 10 points of each other. DEFT II indexes were higher than I or III due to problems inherent in the TP & TRP indexes. Each was twice as large as the others as the multiplicative combination of deficit and difficulty is not contained in DEFT II. When this is compensated for, the DEFT II indexes are the same as I and II. The average disagreement between the raters was about 9 points on a 100 point scale. In the opinion of the authors, these results show excellent interrater agreement.

Rose and Martin suggest some transformation of the efficiency scales to reduce variance. They also suggest that more than one rater apply DEFT and an objective measure of deficit of the trainees be incorporated in DEFT II.

A sound basis is needed to estimate the practice time and number of TD/S and Weapon Systems (WS) needed for training. We found no models that attempt to estimate the number of TD/S vs. WS needed for training in the conceptual phase of WSAP. This is a serious deficiency since these items are likely the most costly in training and represent the most important performance tasks in training and in battle. Further, as Dawdy and Hawley (1983) point out, effective scheduling within the training program frequently revolves around these major items of equipment. By identifying training bottlenecks, their analysis provides information to TD/S designers regarding the usefulness of developing part-task trainers and scheduling TD/S and the WS.

The design and development of a "family" of TD/S and part-task trainers imposes additional performance prediction requirements. The TDDSS Model would, presumably, treat each TD/S in the "family" as separate tasks for which performance and time would be estimated and weighted using the MAUM technique. The CBP method is flexible and might estimate performance and/or time for each TD/S singly or in combination depending on requirements. TRAINVICE and DEFT presumably could estimate transfer in several ways:

1. Sequential transfer from TD/S A to B to C where transfer is from TD/S to TD/S to WS;

2. In various combinations such as:

   TD/S A to WS
   TD/S A & B to WS

TD/S design across a number of courses and for institutional training vs. unit exercises (operator training) or OJT (maintenance personnel) has not been given clear attention.

In empirical studies, Orlansky and String's (1979) evidence shows that in most cases TD/S can be effective substitutes (measured in terms of their transfer effectiveness ratio) for part of the training on the WS, are much less costly (can be amortized within 2 to 4 years) and avoid safety problems that may be attendant to a naive trainee attempting to operate the WS.

Four empirical studies were found. Bickley and Bynum (1980) had instructors estimate the number of trials needed on a TD/S for helicopter flight training vs. the number of trials needed on the helicopter itself for 28 tasks. These figures were used for an empirical trade-off design to determine the number of TD/S trials that would continue to transfer to the WS. Examination of the results showed that the instructors identified the correct trade-off points for only about one-half the tasks. A Navy study (Cicchinelli, 1984) on installing a new training device (electronic equipment maintenance trainer) had instructors indicate how much training on the TD/S seemed to be enough relative to the WS. How this estimate could have been predicted is not readily evident.

Provenmire and Roscoe (1973) empirically analyzed the incremental transfer effectiveness of a ground based general aviation trainer for the successive increments of trainer vs. flight time and found practice trade-off points and substantial cost savings. Boldivici, Bessemer and Haggard (1985) in a validation and verification test report of the M1 Unit-Conduct of Fire Trainer (UCOFT) noted that SMEs underestimated practice time by one-third, compared to the empirical results they obtained. It is not clear how TRAINVICE or DEFT could deal with these problems, except by making specific estimates of time, practice, and number of TD/S and WS needed. Research is needed at both predictive and empirical level to adequately address this problem.

Trade-off analyses also need to be planned for operating and maintenance personnel for OJT, refresher training, skill maintenance and exercises. As far as we can tell, this has not been done in the conceptual phase of the WSAP or after the WS has been fielded.

Adams (1979) raised a number of questions about predicting transfer effectiveness. Among his concerns was whether the TD/S would be used by instructors consistently. This issue is not addressed in the TRAINVICE or DEFT models. Although this is a concern, it can be addressed in part by designing and implementing instructor training to standardize TD/S use or by designing the TD/S to be less reliant on instructor judgments of performance.

The TD/S may itself serve as the performance criterion. Transfer effectiveness to WS performance in first level training may not always be a requisite measure. Some TD/S tasks prepare students for unusual and infrequent battle conditions that would not likely be encountered in training or the field. In these tasks, the TD/S performance and combat exercises may be the closest approximations available to performance criteria.

DISCUSSION AND SUMMARY

Table 4-4 summarizes selected characteristics of the predictive CTEA models. All models have been developed for or applied in the Army. HARDMAN TRRA was developed for the Navy but is being actively adapted to Army requirements.

The models differ in their areas of application. TDDSS and TRRA are used to predict entire courses, with particular emphasis on the conceptual phase of WSAP. However, TDDSS provides performance estimates, a unique method of utility weighting (MAUM) of performance estimates, course time estimates, and, when coupled with TDDA (see Chapter 5) incorporates prescription as well. HARDMAN TRRA, designed to be part of a MPT model for use in the conceptual phase of WSAP, has so far not included performance estimating. TRRA is referred to by its authors as a quasi-training model in that it is not intended to provide a complete design, but to enable trade-offs to be made among hardware, manpower, personnel and training considerations. The HARDMAN authors acknowledge that this approach may be a short-term solution (HARDMAN, Vol. I, undated) to training prediction and prescription. Both models are considered to be evolving and may change in the next few years. Although both models have been used for a number of applications, information about reliability and validity is sketchy. TDDSS reports consensual reliability of three to nine raters. An independent evaluation of the total HARDMAN model is reported, but it is difficult to say how this evaluation applies specifically to TRRA.

Kraft and Farr's (1984) proposed CTEA for CBI in JSEP is a unique model in many respects in that it is the only model found that addresses prediction for CBI, and it is not oriented specifically to new weapon systems (nonsystem). Further, it seems to be addressing the distribution of training across a number of job skills courses of instruction distributed throughout multiple training sites,

Table 4-4

Selected Characteristics of Predictive CTEA Models

Model	Branch of Service	Applications	Type of Training
TDDSS	Army	Entire Course	Institutional, OJT
HARDMAN TRRA	Navy/Army	Entire Course	Institutional
JSEP	Army	Computer Based Instruction	Non-system Multi-course Multi-site
CBP	Army	Training Device	Skill
TRAINVICE	Army	Training Device Transfer	Skill
DEFT	Army	Training Device Transfer	Skill

Model	Performance Estimates	Course Time Estimates	Reliability and Validity	Cost Analysis
TDDSS	Yes	Yes	Consensual Reliability No validity	Yes
HARDMAN TRRA	No	Yes	Independent evaluation of MPT system	Yes
JSEP	Yes	Yes	No-In Development	Yes
CBP	Yes	Yes	Preliminary Validity; No Reliability	Yes
TRAINVICE	Ratings of Transfer	No	Preliminary	No
DEFT	Ratings of Transfer	No	Reliability; No Validity	No

whereas TDDSS and TRRA, from the information available to us, have not addressed these issues. As the JSEP CTEA is still in development, reliability and validity information is not yet available. Chapter 6 provides a further discussion of the need for further development of predictive and prescriptive CTEA models for CBI systems.

CBP, TRAINVICE and DEFT are all concerned with training devices. CBP, however, is radically different in its approach from the TRAINVICE/DEFT models, in that it outlines general strategies and makes predictions based on comparison cases of performance on a specific device, course time requirements and costs. Validity studies are reported as in progress, but were unavailable for first-hand examination. Reliability and validity are a particular concern as the generality of the approach and the SMEs employed to make judgments could lead to widely differing results. CBP, for the situations to which it applies, may, however, be advantageous, compared to TRAINVICE/DEFT, in obtaining time, performance and cost estimates for specific devices and in the limited demands made of SMEs reportedly about one hour per device.

TRAINVICE and DEFT models provide raters' estimates of transfer of training. Grounded in transfer of training research and theory, ratings are made on many dimensions and combined by complex formulas. The models do not provide estimates of course time requirements or cost analysis. Conceivably, these models could be supplemented by asking SMEs to provide course time requirements, costs and specific estimates of performance such as hits on target. DEFT makes an important contribution in providing three levels of analysis for varying information inputs.

A number of reliability and validity studies of TRAINVICE have been undertaken including the comparative validity study of the four TRAINVICE models, demonstrating that two of the models are promising. However, with the reconceptualization of TRAINVICE into DEFT, further reliability and validity studies are needed. Time and effort demands on SMEs are not clearly reported but appear to be in the range of one to three days per device. It has yet to be demonstrated empirically that DEFT is an improvement over TRAINVICE. Further, the ability of these models to discriminate among "good" and "poor" training device designs is open to question. It is our understanding that these issues are being addressed by studies now underway (Mirabella, Personal Communication, July, 1985). Also, as noted earlier, trade-offs between time on a TD/S and a WS need further empirical and predictive research.

The references in Table 4-4 to reliability and validity should be interpreted cautiously. Reliability and validity methods are discussed in detail in Chapter 8. It remains questionable whether any of the models, TRAINVICE excepted, have used clear and stringent methods of obtaining reliability and validity estimates. In addition, cost analysis does not necessarily mean that an adequate model has been developed.

Information input requirements vary to some extent for each of the models, but generally require, as a minimum, a functional description of the weapon system, job analysis and task analysis. TDDSS and TRRA develop a job or task analysis as part of the process. CBP, TRAINVICE, and DEFT assume the availability of training device functional descriptions. CBP assumes the availability of similar training devices or simulators. TRAINVICE assumes a task analysis detailed into subtasks or skill elements. The innovation in DEFT is the provision of three levels of analysis depending on the detail of the task analysis, making it more responsive than TRAINVICE to early developments in a training device.

In summary, there are a number of predictive models that can be adapted to a variety of uses. Each model is undergoing further development. Although each model has been applied in a number of settings (JSEP excepted) there are many questions remaining regarding the predictive issues they address, information yield, reliability, validity and ease of use. Further developments are also needed in the development of predictive CTEA cost models (see Volume II), models for CBI systems and models for training that involves multiple courses of instruction and instruction at a variety of sites (non-system, exportable models). Other issues are addressed in Chapters 7 and 8.

Chapter 5

PRESCRIPTIVE COST TRAINING EFFECTIVENESS ANALYSIS (CTEA) MODELS

INTRODUCTION

Many prescriptive models are available for cost and effectiveness analysis of training programs. Prescriptive analysis incorporates job and task analysis, selection of tasks for training, clustering and sequencing of tasks in the program, recommendations for methods and media for delivery, determining the setting, and estimating resources. Some methods include cost analysis. The goal of these models is to select among training alternatives and aid the analyst in choosing a program which is the most cost effective in achieving the objectives of the training program.

The overall process is captured in the process flow chart shown in Figure 5-1. Emphasis is on the design of media, methods, delivery systems and TD/S for entire institutional training programs early in the Weapon Systems Acquisition Process (WSAP) All of the models yield "figures of merit" based on the qualitative media/method characteristics.

These models are reviewed only briefly in this report. For detailed reviews see Rose et al. (1980), Matlick et al. (1980) and Knerr et al. (1984). Models described in this chapter are:

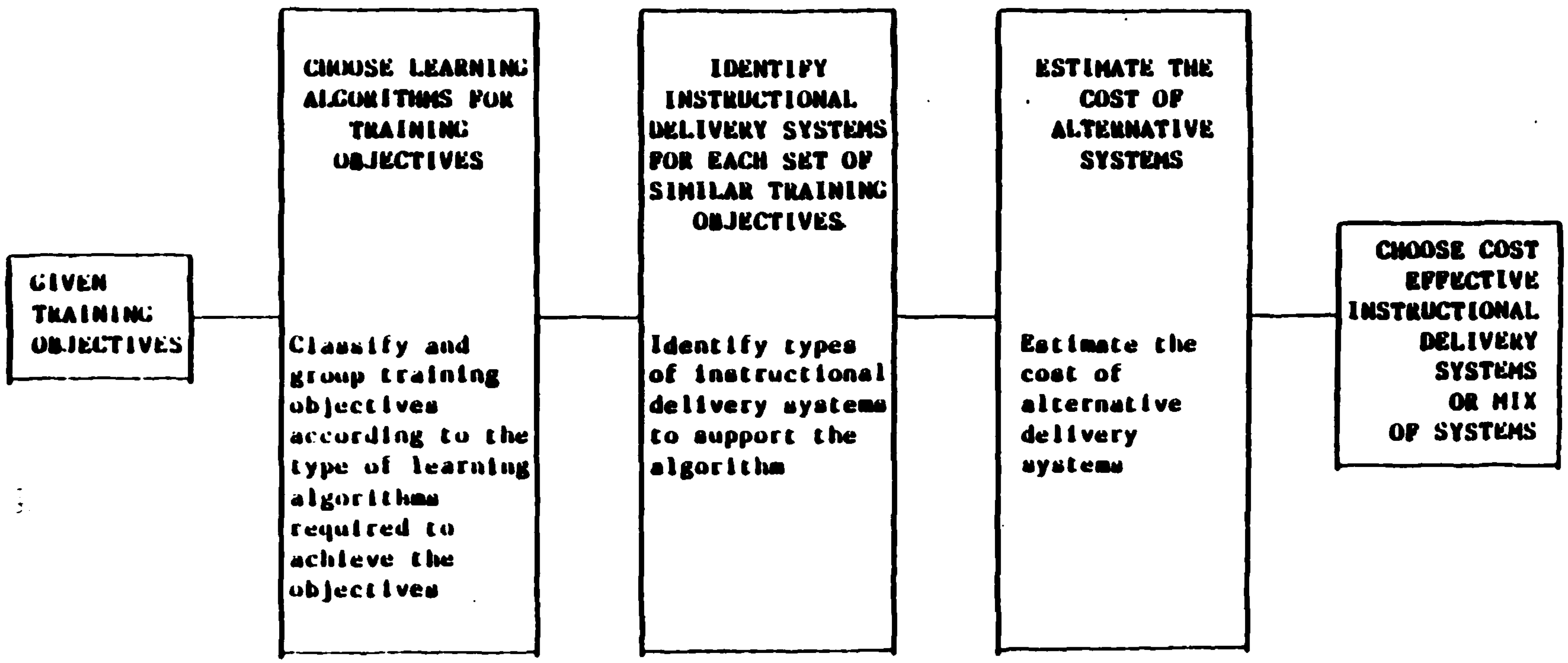

Figure 5.1 Process Flow in the TECEP Technique*

*From Braby, Henry, Parrish, and Swope, 1975, p. 13.

1. Training Efficiency Estimation Model (TEEM)

2. Training Consonance Analysis (TCA)

3. Training Development Decision Aid (TDDA)

4. Method of Designing Instructional Alternatives (MODIA)

5. Training Requirements Analysis Model (TRAMOD)

6. Training and Cost Effectiveness Prediction (TECEP)

7. Training Device Design Guide (TDDG)

8. Taxonomy of Device Features

The last three models are specifically training device specification models or guides. They are prescriptive and make recommendations concerning specific features to be incorporated within a device during the design and procurement phase.

TRAINING EFFICIENCY ESTIMATION MODEL - TEEM

Jorgensen & Hoffer's (1978) TEEM is a prescriptive method which generates a cost/efficiency ratio for methods and media mixes for training programs. The efficiency metric is an index of match between the learning needs and methods/media that meet these needs. TEEM procedures incorporate task description and analysis, training program generation, analysis of costs and comparison of cost and efficiency.

TEEM inputs include a task list and knowledge of the weapon system. Tasks are then rated on the following variables: stimulus characteristics, response characteristics, information feedback and functional context variables. Tasks are clustered together according to the communality among the variables; pre-coded media and methods are then matched with the tasks. The media/methods with the highest number of matches across these variables are selected as the most efficient. An efficiency ratio is generated with the score of the estimated program in the numerator and the efficiency score of a base program in the denominator. The efficiency ratios range from 0 to 1. The original TEEM program used an ideal program (no cost and resources constraints) as the base program. The new TEEM program, however, uses actual equipment as the base case for comparison. TEEM is automated and implemented on the Hewlett Packard 21MX and Apple II microprocessor. A user's manual has been prepared for its use.

TRAINING CONSONANCE ANALYSIS - (TCA)

Hawley & Thomason's (1978 a & b) TCA is a modification of TEEM and is designed to predict training effectiveness. TCA uses the same variables as TEEM for task descriptions and media-method combinations. TCA describes the consonance of media-method combinations and the task descriptions in terms of "training deficiency," "training excess" and "training redundancy." Then diagnostics help differentiate between the efficiencies of alternative training programs. An efficient training program would have only a small number of excesses and redundancies. The task descriptions are compared to the media-method descriptions variable by variable. When a variable is common to both, a training consonance is scored; otherwise training deficiency, excess, or redundancy is scored. The training consonance is computed by dividing the total number of training consonances for the task by total number of variables in the task description. The scores are aggregated over groups of tasks and over all tasks. Outputs are given for each task, groups of tasks and all tasks. Outputs include training consonance ratios, training deficiencies, excesses and redundancies for each task. Outputs for task groups and all tasks include total number of training consonances, total number of variables in task descriptions, and the training consonance ratio.

TRAINING DEVELOPMENT DECISION AID - TDDA

Hawley and Frederickson (1984) designed an automated model for training design and evaluation which includes cost analysis. This is an early version of the prescriptive part of the Training Development Support System (TDDSS). It contains a data base, a model base, and a software system that includes data base and model base management and dialog generation and management. It is based on ISD methods.

The training design processes have three phases. The first phase includes function analysis to generate a task list and analyze task criticality to select tasks for training. These tasks are then rated on ten dimensions and classified into three levels of training. The second phase is task analysis in which performance standards, conditions, skills/knowledge and task consequences are generated. A task hierarchy is generated for tasks that require training.

In the third phase, learning requirements analysis identifies the special features of the task, functional learning requirements, training constraints, and resources. The functional learning analysis produces information on which media and devices would be appropriate for the task and how to prepare the training content. The analysis emphasis is on the characteristics of the learner.

Training evaluation has the following components:

1. Resource projection model
2. Cost estimation model
3. Benefits analysis model
4. Trainability analysis model
5. Cost-benefit integration

In the benefits analysis, priorities for the training modules, evaluation dimensions on context variables with ratings of merit for each training program alternative, are combined to compute a figure of merit for each alternative. It utilizes a general Multi-Attribute Utility Measurement (MAUM) method described in detail in Chapter 4.

METHOD OF DESIGNING INSTRUCTIONAL ALTERNATIVES - MODIA

MODIA was developed by Carpenter and Huffman in 1977 to help the Air Force manage its resources for formal training "... by relating the requirement for training resources to course design and operation" (Knerr et al., 1984,). MODIA considers alternative training designs and applies primarily to the design and development stages of ISD. MODIA has four major components:

1. Description of options for course design

2. User interface (UI)

3. Resource utilization Model (RUM)

4. Cost model (MODCOM)

The first component identifies course designs and the data required to use them and aids in making design choices. The second component (UI) aids the user in describing the various factors in a course which results in a course design. The third component (RUM) analyzes the course by simulating the flow of students through the course in various combinations. The fourth component (MODCOM) has five cost outputs based on the analysis of manpower and other training expenses.

MODIA has been successfully applied to the Air Training Command's Flight Facilities Equipment Repairmen course. Kessler's School of Applied Aerospace Sciences evaluated the utility and accuracy of MODIA. Their results supported MODIA's utility and accuracy and also found that for three of the five courses, the alternatives generated by using MODIA were less expensive than baseline courses. The MODCOM step of MODIA predicts costs of a course and the RUM step estimates its operational effectiveness, but MODIA does not predict the instructional effectiveness of a course.

TRAINING REQUIREMENTS ANALYSIS MODEL - TRAMOD

TRAMOD (Czuchry et al., 1978) aids in making decisions regarding which tasks should be trained and the efficacy of training techniques in meeting the training objectives. This is a computerized model for use early in the WSAP. TRAMOD guides design decisions by comparing training alternatives and making cost-related training decisions. The first TRAMOD component, Task Block Generator, tests the training characteristics of task information supplied by the user. The training characteristics are quantified in terms of five parameters: level, learning difficulty, criticality, frequency and collating tasks which are best trained as a block. The second component, Training Plan Generator, requires data inputs such as time limit in months, dollar cost expense limit and requirements. The program recommends training mode, media and method of instruction. After the user specifies constraints on the training plan, the third component, Training Program Generator, yields a detailed schedule of the training program which specifies task blocks to be trained, training methods and media, training modes, and training time.

TRAINING AND COST EFFECTIVENESS PREDICTION - TECEP

TECEP (Braby, 1973) is a synthesis of techniques for choosing instructional media and prescribing training programs during the system development and acquisition cycle. This model was generated by the Navy's Training Analysis and Evaluation Group (TAEG).

The TECEP objective is to match training objectives with an appropriate cost-effective training system in an analysis of the type of learning algorithm associated with each training objective. The following steps are taken in TECEP.

1. **Classify training objectives.** There are twelve classes of tasks. The training objectives are classified, and TECEP provides a learning algorithm

for each task category. A learning algorithm is a step-by-step procedure for learning the tasks in that category to meet the training objectives.

2. <u>Identify instructional delivery systems.</u> Special instructional delivery systems are identified for each task category and those which are the most capable of supporting learning are selected. Practicality of the media are kept in mind.

3. <u>Cost analysis.</u> TECEP has automated capabilities for generating cost estimates for procedures, implementation, etc., for the alternative training delivery systems.

TECEP has been incorporated into the ISD model. It has been compared to nine other media selection models and it received the highest rating by the staff members of TAEG. This model is based on learning theory and research on the applications of guidelines for designing training systems.

TRAINING DEVICE DESIGN GUIDE - TDDG

TDDG (Smode, 1973) focuses on specifying human factors input to training device design after the device functional requirements have been specified. Human factor inputs include information requirements needed during instruction for the trainee and the instructor stations, and setting limits of fidelity. Techniques implementing this model are outlined. The model is not automated and has few quantified indices. It covers performance measurement.

TAXONOMY OF DEVICE FEATURES

Hughes (1978, 1979) provides a classification of training device features into enabling and instructional features. Enabling features are further divided into equipment (display, controls, etc.) and environmental (audio, visual simulations) subclasses. Instructional features are divided into active and passive subclasses.

These are assessed, not by fidelity, but by their impact on improvements on trainee performance and instructor function. Active subclass includes features which have direct contact with the trainee; passive subclass is comprised of features which have no direct contact with the trainee.

DISCUSSION AND SUMMARY

Table 5-1 shows selected characteristics of these models. The emphasis of six of the models is on the design of media, methods, delivery systems and TD/S for entire institutional training programs early in WSAP. Several models may also be used to improve training in later phases of WSAP (MODIA, TECEP, TDDA). Two guides (TDDG, TDF) and the other models apply to training device design and ultimately might be complementary to the TRAINVICE and DEFT models reviewed in Chapter 4. MODIA concentrates on training resource allocation and combines with TRAMOD for prescription in the Air Force.

Among the course design prescriptive models developed for the Army, TDDA when combined with TDDSS fulfills all of the criteria of providing for performance estimates, time estimates, and cost analysis. TDDA/TDDSS, in contrast with the course design models, does not base its effectiveness measures on "figures of merit" derived from the media and methods. Instead, TDDA/TDDSS provides documentation for training developers to guide the training development process. The match between prescribed and developed media and methods would provide the measure of predictive validity. The effectiveness measures found in the companion

Table 5-1

Selected Characteristics of Prescriptive CTEA Models

Model	Branch of Service	Applications	Type of Training	Performance Estimates
TEEM	Army	MMD*, Training Device Design	Institutional Skill	No
TCA	Army	MMD, Course, Training Device Design	Institutional Skill	No
TDDA/ TDDSS	Army	MMD, Training Device Specifications, Tasks	Institutional, OJT, Other Skill	In TDDSS
MODIA	Air Force	Resource Allocation	Institutional Skill	No
TRAMOD	Air Force	MMD, Training Device Design	Institutional Skill	No
TECEP	Navy, Army	MMD, Training Device Design	Institutional Skill	No
TDDG	Navy	Training Device Design	Skill	Transfer
TDF	Army	Training Device Design	Skill	No

Model	Course Time Estimates	Reliability and Validity	Cost Analysis
TEEM	No	Not Found	Yes
TCA	No	Not Found	Yes
TDDA	In TDDSS	No	Yes
MODIA	?	Preliminary	Yes
TRAMOD	Yes	Not Found	Yes
TECEP	No	Preliminary	Yes
TDDG	--	No	No
TDF	No	No	No

* Media, Methods, Delivery Systems

TDDSS model are based on weighted predictions of actual performance relevant to the course of instruction (Hawley and Frederickson, 1983). In contrast, the "figures of merit" based on media and methods seem to make many unverifiable behavioral and mathematical assumptions as estimates of effectiveness. Further, it may prove difficult to validate them in terms of discrepancies between prescribed and fielded training designs.

HARDMAN TRRA, reviewed in Chapters 3 and 4 (not shown in Table 5-1), is also partially prescriptive but is a "quasi-training design" focused upon trade-offs between hardware, manpower, personnel and training. It does not at this time have the advantages of the other detailed prescriptive models.

Information yield, reliability, validity, and ease of use need to be better established particularly for the Army models. Only MODIA and TECEP report general evaluations. The training device design guides, still in their formative stages, will have to address reliability, validity and integration with transfer of training models.

With two exceptions we found no references to designing training across the training spectrum to achieve better effectiveness. TRRA reports (HARDMAN, Vol. I, undated) designing training for operator and maintenance personnel. TDDA (Dawdy and Hawley, 1983) mentions considering the division of tasks between institutional training and OJT. MODIA and TRAMOD confine themselves to institutional training. Non-system, multi-site, multi-courses of instruction were not apparent.

There was insufficient information available to tell how well these models addressed student throughput or instructional management issues. Further, cost analysis does not necessarily imply a sound cost model.

In summary, there are a number of prescriptive models for entire courses, tasks and TD/S design available to the Army to design and improve training at various stages of the WSAP. With refinements these models can probably be used for course design across the training spectrum.

There are as yet many unresolved issued regarding these models. Training effectiveness issues are discussed in Chapters 7 and 8.

CHAPTER 6

COMPUTER-BASED INSTRUCTION (CBI) AND COMPUTERIZED ADAPTIVE TESTING (CAT)

INTRODUCTION:

This chapter discusses applications of computers in the military for training and testing. The cost-effectiveness of these methods, however, has not been fully assessed and they are the subject of some controversy.

As with computer technology in general, CBI and CAT are evolving methodologies. At this stage of their development they may be considered promising techniques still in need of empirical study before sound predictive and prescriptive models can be developed and advantages or disadvantages compared to other delivery and distribution systems clearly discerned. Furthermore, CBI represents one type of non-system training.

For these reasons, this chapter discusses available models and empirical studies for CBI and CAT. CBI is first defined. Models and studies discussed are:

1. Kraft and Farr's (1984) predictive CTEA for CBI
2. Suitability of other available predictive and prescriptive models for CBI
3. Seidel and Wagner's (1977) Cost-Effectiveness Specification for Computer-Based Training Systems
4. Orlansky and String's (1979) review of comparative empirical studies of CBI vs. conventional methods of instruction

The discussion of CAT focuses upon the CAT concept, empirical studies and expected developments.

COMPUTER-BASED INSTRUCTION

Definitions

Computers are capable of performinga number of functions in the instructional setting. Although there is overlap and combination of functions for a given system configuration, for convenience in conveying major purposes of a system in relation to the courses of instruction which it supports, CBI may be categorized in three ways:

1. Computer-assisted instruction (CAI). All instructional materials, e.g., lessons and tests, are stored in the computer. The student interacts with this material in real time via a terminal and display system. The computer performs many functions, such as diagnose student performance, provide prompts, prescribe lessons, maintain records of student progress, and predict individual course completion dates. The instructor plays a tutorial role in this CAI mode.

2. CAI "Drill and Practice". An alternate CAI mode is to provide instructor support by means of "drill and practice" in which the main body of instructional content is presented by the instructor. The computer is used to provide review, additional problems, practice, individual remediation and tests. In this case, the computer assumes the role of a tutor.

3. Computer-managed instruction (CMI). Instruction takes place away from the computer. The computer scores the tests and interprets results to each student; advises him to take following or alternative lessons; recommends remediation; and manages student records, instructional resources, and administrative data. In some cases, only testing, scheduling and record keeping are carried out by the computer.

The definitions given by the military have only distinguished between CAI and CMI. The CAI "Drill and Practice" distinction is important as it impacts the instructor role vis-a-vis the computer, reportedly makes a difference in the courseware development effort, and may require differing facilities and locations of the computer terminals.

Additional distinctions are also potentially useful for predictive and prescriptive CTEA. Some CAI systems couple other media (such as audio tapes, slide-tapes, or VideoDiscs) to them to provide audio or realistic motion capabilities not easily provided by the currently available computer technology. In a CMI mode, the student is sometimes directed to study lessons contained on these media or to manuals, texts or group instruction. In all cases, instruction is individualized or self-paced to a greater extent than in conventional classroom training.

In a complete course of instruction for military skill training, CBI can only play a partial role as it is best suited to instruction of knowledge, information, procedures, and diagnostics. Computer-embedded instruction is a closely related concept to CBI, but still distinctive. In computer-embedded instruction, lessons and diagnostics are contained on a computer to aid soldiers in handling computer-related work problems. The computer-embedded instruction concept will not be considered further in this report as no information was available. It is, however, worth additional investigation.

KRAFT AND FARR'S (1984) PREDICTIVE CTEA FOR CBI IN THE JOB SKILLS EDUCATION PROGRAM (JSEP)

This model was reviewed in detail in Chapter 4; however, a number of observations are worth stating.

1. The JSEP Model is presently under development. No data are yet available. The model proposes to use methods of decision-making under uncertainty to test various mixes of terminals, courses, students, and instructors.

2. The application appears to be one that involves multiple sites and multiple courses.

3. The system configurations, types of CBI and the course tasks were not specified. Presumably, courseware development relevant to a prescriptive CTEA is being undertaken as a separate task.

4. No mention is made of comparative analysis to other instructional delivery systems. Presumably, a commitment has been made to use CBI.

5. The model is not related by the authors to the early phases of the WSAP. The emphasis appears to be on furthering the exportability of the CBI technology to multi-site, multi-course applications (non-system training).

The results of the JSEP Model may prove useful in gaining an understanding of some of the dimensions relevant to a predictive CTEA for CBI systems.

SUITABILITY OF OTHER AVAILABLE PREDICTIVE AND PRESCRIPTIVE CTEA MODELS TO CBI

A review of the models presented in Chapters 4 and 5 was undertaken to try to determine whether certain concepts and methods might be appropriate to further development of predictive and prescriptive CTEA for CBI systems. This review produced somewhat limited results as we had access to only a number of the complete modeling documents. The review was based, therefore, on descriptions of the models contained in Knerr et al. (1984) and Matlick et al., (1980). Nonetheless, a number of observations can be made, subject to a first-hand examination of all the models for adaptability to CBI systems:

1. TDDA/TDDS, TEEM and TECEP appear to be adaptable to including CBI as delivery systems alternatives, along with other media and methods.

2. In the applications described in various reports, no mention was made of adoption of CBI systems. Interviews with developers and users and examination of the prescribed training plans could, however, show otherwise.

3. The prescriptive learning algorithms (stimulus, response, still, motion, etc.) may be promising if appropriately restructured to adequately reflect different types of CBI system configuration capabilities.

4. Performance and time estimation are equally relevant to prediction in CBI as in other delivery systems.

5. The Multi-Attribute Utility Method (MAUM) or a similar method may be suitable for assessing the "worth" of various CBI system characteristics in comparison to other delivery system configurations. These characteristics might include: (a) performance advantage, (b) learning time advantage, (c) stimulus-response control, (d) instructor availability, load and role changes, (e) testing and scoring convenience, (f) scheduling convenience, (g) administrative convenience, (h) individualization, (i) remediation, (j) immediacy of feedback and knowledge of results, (k) difficulty vs. ease of course revision and updating, (l) portability and distribution, (m) student interest and motivation, (n) student fatigue, (o) computer experience, and (p) enhanced performance (transfer) on subsequent knowledge and skill tasks. Measures would need to be developed appropriate to differing CBI modes, other delivery systems and various audiences (students, instructors, administrators, training designers, and training developers).

The emphasis of the existing models on formulating training programs for emerging weapon systems appears to place constraints on consideration of CBI systems.

1. Emphasis in current models is oriented toward entire training programs, or training devices (TD/S) in an institutional, system specific

setting. CAI is most suitable for knowledge and information instruction across a variety of courses and sites; CMI to managing, scheduling and testing where individualization or remediation is an important objective.

2. Current models emphasize using available teaching resources. CBI systems are not widely available at present.

3. Lead times for acquiring CBI systems and developing courseware can range from three to seven years, far too long to be useful in initial fielding of training for a new WS (Hawley, Personal Communication, November 1985).

4. Performance deficiencies and training management problems for which CBI may be considered a viable alternative may not be easily recognized in the early phases of training and WS development.

5. Emphasis of current models on cost minimization as opposed to performance maximization, and the uncertainties of CBI effectiveness, add additional constraints to its consideration early in WSAP.

In time, however, early prediction and prescription of CBI as part of a training plan may become possible. This will likely depend on CBI systems becoming more available; courseware development capability improvements resulting in reduced development lead times; task analytic improvements providing early detection of prospective performance deficiencies; the real benefits and deficits of CBI being more clearly established by research; and a continuing decline in the cost of computer systems. The capabilities of currently available computers are probably more than sufficient for their foreseeable applications to instruction. Areas that need development are related to how best to program and use CBI for instructional purposes.

SEIDEL AND WAGNER'S (1979) COST-EFFECTIVENESS SPECIFICATION FOR COMPUTER-BASED TRAINING SYSTEMS

This methodology was developed by Seidel and Wagner while at the Human Resources Research Organization for the Defense Advanced Research Projects Agency. It was designed to facilitate the purchase, monitoring and evaluation of CBI training systems. The method provides a standardized structure through which CBI system effectiveness and costs can be derived. Training effectiveness measures include performance achievement measures and time measures for within-course and end-of-course criteria. Attrition rates, instructor ratings and attitude scales are also discussed. Suggestions are made to the training developer for weighting effectiveness measures. This emphasis on student performance and time envisioned a CAI mode rather than a CMI mode in which instructional management and student record-keeping would be paramount concerns. This method was developed as a guide for empirical studies but may provide useful guidance in developing a predictive CTEA. There is little guidance provided for prescriptive design. The cost effectiveness procedures are discussed further in Volume II.

ORLANSKY AND STRING'S (1979) REVIEW OF COMPARATIVE EMPIRICAL STUDIES OF CBI

Orlansky and String reviewed about 30 studies conducted since 1968 which evaluate CBI versus conventional training and individualized instruction (programmed instruction) in the Military. They defined conventional and individualized instruction as follows:

1. Conventional instruction, where an instructor may use lectures, discussions, laboratory demonstrations, and tutorial sessions. Groups of students proceed through the curriculum at the same pace; differences in achievement among students are reflected in grades at the end of the course.

2. Individualized instruction, where each student proceeds at his own pace through the curriculum that is arranged in a series of lessons and tests. Mastery of each lesson is set as a condition of progress. Differences among students are reflected in the amounts of time needed to complete the course, although grades may also be given. In general, an effort is made to assure about the same level of achievement for all students.

Several criteria were used to estimate training effectiveness.

1. Student achievement - All studies used student achievement in school as a measure of effectiveness. No important differences of practical significance in student achievement were observed between the various methods of instruction. Student achievement was measured by performance on tests administered at the schools. A total of 40 comparisons was made. In 24 cases, student achievement with CAI was the same as with conventional instruction, superior in 15, and inferior on one. CMI was compared to conventional instruction in 8 cases. No significant differences were found in student achievement. In five cases of comparison of CBI to programmed instruction, achievement was the same in 4 cases, and CAI was superior in 1. Overall, although the differences found in these studies were statistically significant, they were judged not to be large enough to have practical significance.

2. Student time savings - CBI students saved about 30% (median value) time over conventional instruction. There is, however wide variation in time-savings reported by the various experiments. On the whole, 40 comparisons showed a range of -31% to +89% in time savings with CBI as compared to conventional training. In three cases, student training time was higher as compared to conventional instruction. Another study showed no effect on time savings.

Programmed instruction without computer support showed student savings similar to CAI or CMI.

3. Student Attrition - Student attrition rates increased for CMI in Army and Navy studies. No increase in student attrition was reported in CAI. Marginally meaningful data on student attrition were gathered from the Air Force Instructional System (AIS), where four courses were implemented on a CMI system over a period of 4 years, and by the Navy CMI system, where data from seven courses were available. Student attrition rates increased in all 4 courses on AIS. However, academic attrition rose in all courses at Lowry AFB over the same period. In the Navy CMI system the attrition rate increased from 3.2 to 4.6 percent after the implementation of CMI in the seven courses. Four experimental programs yield data on student attrition. One found attrition 22% lower for the CAI group, and 3 others found no difference.

4. Attitudes of Students & Instructors - Few studies reported the attitudes of students and instructors towards CBI. While students tended to prefer CBI to conventional instruction, the instructors held an unfavorable attitude towards CBI versus conventional instruction. Data show that of the 39 experiments which addressed student attitudes, 37 studies showed student attitudes to be favorable to CBI, one showed no difference, and one showed unfavorable attitudes towards CBI. Instructor attitudes were reported by only 9 studies, 8 of which found instructor attitude to be unfavorable, and 1 found it to be favorable towards CAI.

Orlansky and String (1979) note that the effectiveness of CAI and CMI could not be compared directly as none of the experiments gave adequate information about the instruction being compared. They stress the importance of measuring effectiveness by gathering data on performance in the operational setting. None of the studies reported such data.

Only 8 studies reviewed by Orlansky and String (19[illegible] addressed cost, and cost data described were part [illegible] expenditures incurred during the course of the [illegible] one case, CAI was judged to be not cost [illegible]

System). The Navy CMI claims to have saved $10 million in FY 1977, and the Air Force AIS claims savings of $3 million in 1978. These figures were based on student time savings translated into dollars. The costs of providing CMI were not considered. They recommend a number of improvements to provide complete rather than selected costs to evaluate CBI.

Individualized instruction (programmed instruction) was treated as a special case by Orlansky and String (1979). They found a number of studies in which programmed instruction was an intermediate step towards computerization of conventional instruction. Programmed instruction is based on the same learning and training principles utilized in CBI but without computer support. Programmed instruction of five courses saved 64% student time over conventional instruction. The CAI version saved 51% for seven other courses. They point out that the issue of whether computer support to individualized instruction would lower costs by reducing the number of personnel and instructors, by maintaining student records, and by modifying and updating courses, has not yet been explored.

The review of empirical studies of CBI vs. conventional instruction is being updated by Orlansky and String (Personal Communication, August, 1985). Newer studies might address the issues mentioned above.

Discussion of CBI

There are merits to using computer-based instruction in the military. The results of the experimental programs comparing CBI to conventional instruction are clear, i.e., CAI produces about the same student achievement at school as does conventional instruction and saves student time. Costs are not reported by enough studies to draw firm conclusions.

However, one major difficulty is in attempting to neatly categorize various types of instructional methods as either conventional or computerized. There can be various system configurations and baseline training configurations with different cost effectiveness implications. The system configuration would depend on the combination of hardware, software and other facilities and its range of applications. A system having a wide range of applications may be more costly to acquire, but may also prove to be more cost effective.

Different features of CBI instructional strategy may well have specific training applications which may be at the course level, task level, subtask level, or skill level. This, too, may be narrow or broad in its application. These applications would determine the configuration of the CBI system and thus affect instructor-equipment-student mix and related cost and effectiveness measures.

Comparative analyses of CBI training have been done, mostly using "conventional" instructor-led group instruction as the baseline. It is these studies that show a learning time advantage for CBI, taken over a large number of studies. This type of comparison represents the current "state-of-art" and is of some value in suggesting the relative efficacy of CBI and "conventional" instruction. However, it is a limited approach for a number of reasons:

1. "Conventional" instruction is ill-defined. It is often assumed to be a lecture or lecture-discussion approach. While this may be partly accurate, it is incomplete in that it does not describe the methods or media employed by various instructors, differences in teaching effectiveness of instructors or variation in student abilities and motivation.

2. Courseware design may reconfigure the teaching task to that which can be most advantageously presented via a computer. In the process, there may be a close similarity in presentation and overlap of content presented or they may be quite divergent with the computer adding or subtracting instructional features.

3. Individualization of instruction, one of the major claimed advantages of CBI may, be accomplished in a number of ways that may not need a computer. Included among these are pretesting; pretesting with lessons or assignments assigned based on pre-test results; testing-remediation provisions; reading assignments before a course to bring students to a common level of knowledge; "learning packages" that provide pretest, assign lessons/learning activities accordingly and post-tests when the lesson or learning activity has has been completed; programmed instruction used as the primary presenter of training, for drill and practice or self study; correspondence study; tutoring; media coupled with workbooks or response forms or self-tests such as videodisc, videotape, audiotape, slide-tapes, etc.

Comparisons of CAI and programmed instruction show no differences in study time and performance during acquisition learning. However, costs for programmed instruction were much lower than CAI (Orlansky 1985). Comparisons of CBI with other methods of individualizing instruction have not yet appeared in the military literature. However, they need to be considered further if the relative advantages and disadvantages of CAI and other instructional methods and media are to be better understood.

Despite these difficulties, it is important to note that application of new technological innovations in training and testing is important in the military and certainly worth pursuing. The use of faster computers, with improved memory capacity and response time, has reduced computer and processing costs. The military currently utilizes minicomputers and microcomputers and their availability is expected to increase.

A sound front end analysis, learning algorithm, and instructional management paradigm embodied in a predictive and prescriptive CTEA would likely be helpful in design and also suggest better hypotheses for empirical testing of CBI systems vs. other delivery and distribution systems. It appears to be a propitious time to develop such a model.

COMPUTERIZED ADAPTIVE TESTING (CAT)

Another emerging aspect of usage of computers in the Military is Computerized Adaptive Testing (Rosen, 1985) and Flexilevel Adaptive Testing (Hansen, Ross & Harris, 1977).

CAT is an automated test administration and scoring system, using a computer, whose purposes, compared to conventional paper-pencil testing, are:

1. To discriminate the examinee's aptitudes and achievements at all levels through flexible binary selection of items presented for response. Medium difficulty questions are administered at the beginning of the test and, depending upon the response, less difficult or more difficult questions are presented, capturing the aptitude or achievement levels of examinees more efficiently, i.e., with fewer questions and in less time than required by conventional tests.

2. To reduce the test scoring and administration errors involved in manual testing.

3. To reduce testing time by administering fewer items and still achieve the same degree of

validity and reliability as regular paper-pencil tests.

4. To save publishing costs and development of new forms costs. New items can continuously be developed and updated.

5. To reduce the possibility of compromising the tests, i.e., of students learning about the test items and receiving inflated scores.

The Department of Defense (DOD) has undertaken a CAT-ASVAB project to develop a system to administer the Armed Services Vocational Aptitude Battery which is used to evaluate potential enlisted personnel and make decisions regarding the selection and placement of applicants to military service. Currently the ASVAB is used to test approximately one million applicants per year for enlistment and testing of high school students in the DOD student testing program.

The research on CAT-ASVAB is ongoing. It is the first major military effort in adaptive/tailored testing and is important in terms of serving as a model for analyzing organizational concerns and activities associated with the development of such a system. The costs of implementing and operating such a system have not yet been evaluated.

A study conducted on Flexilevel adaptive achievement testing versus conventional testing showed an 18.4% savings in time for the adaptive test. Nine fewer items than conventional testing were required to obtain flexilevel scores. The flexilevel scores and conventional scores correlated highly, R = .940 (Hansen, Ross & Harris, 1977).

Applications to training related achievement testing are only now being considered. In the future, the CAT concept could aid in providing reduced testing time and improved administrative and scoring accuracy with little of any reduction in test reliability and validity. The CAT concept could be integrated with CBI or stand alone as a separate system.

Chapter 7

ADDITIONAL ISSUES OF TRAINING EFFECTIVENESS ANALYSIS IN CTEA MODELS

INTRODUCTION

This chapter starts with a review of findings and observations regarding the predictive and prescriptive models discussed in Chapters 4, 5 and 6 and a list of variables and constructs abstracted from this literature. Additional issues relevant to training effectiveness analysis are then discussed in relation to the models and the list of variables and constructs. These additional issues include the following:

1. Applications and types of training
2. Performance and time estimates
3. Student throughput
4. Instructional management
5. Motivation
6. Responsiveness of models to policy changes

REVIEW OF FINDINGS REGARDING PREDICTIVE AND PRESCRIPTIVE MODELS

Table 7-1 shows a consolidation of selected characteristics of predictive and prescriptive models taken from Tables 4-3 and 5-1. This table is confined to models applied in Army settings. From this table and the discussion in Chapters 4, 5 and 6, the following observations can be made:

1. Only one set of models, TDDSS and TDDA, combines prescription with prediction of performance and time. This pair of models has also been used for analysis at a task level, trainability analysis, design of OJT, specifications for TD/S, sequencing, and analysis of bottlenecks in course design.

2. HARDMAN TRRA obtains time estimates but not performance estimates and provides a "quasi" training plan for hardware, manpower, personnel and training tradeoffs.

3. Four models, TEEM, TCA, TDDA, and TECEP have been used by the ARMY for institutional and TD/S training design. How useful these models have been for TD/S design is not clear.

4. Three predictive models have been developed for TD/S. The CBP estimates performance, time and costs, and TRAINVICE and DEFT predict transfer of training. Further development and integration of modeling efforts are needed to: (a)predict TD/S time and performance vis-a-vis the WS; (b) integrate prescription and prediction; (c) determine the role of part-task trainers and a family of TD/S in prediction; (d) define analyses appropriate to certain TD/S as "criterion" measures of performance (that is, when TD/S tasks simulate conditions not often encountered); and (e) determine uses of TD/S outside an instructional training setting.

5. Only one predictive model, Kraft and Farr's JSEP (1984), has been developed specifically for CBI systems and exportability across a number of sites (non-system training). Results are not yet available as this model is still in development. As noted in Chapter 6, it is not clear whether existing prescriptive models are well suited to CBI design. Further development is needed in the areas represented by this model, that is: CBI prediction and prescription; exportability; and mixes of training equipment, students and instructors.

Table 7-1

Selected Characteristics of Predictive and Prescriptive CTEA Models Used by the Army

Model	Applications	Type of Training	Performance Estimates	Course Time Estimates
PREDICTIVE				
TDDSS	Entire Course Tasks	Institutional OJT	Yes	Yes
HARDMAN TRRA	Entire Course	Institutional	No	Yes
JSEP	Computer-Based Instruction	Non-system Multi-course Multi-site	Yes	Yes
CBP	Training Device	Skill	Yes	Yes
TRAINVICE	Training Device Transfer	Skill	Ratings of Transfer	No
DEFT	Training Device Transfer	Skill	Ratings of Transfer	No
PRESCRIPTIVE				
TEEM	MMD*, Training Device Design	Institutional Skill	No	No
TCA	MMD, Course, Training Device Design	Institutional Skill	No	No
TDDA	MMD, Training Device Specification	Institutional, OJT, Other Skill	In TDDSS	In TDDSS
TECEP	MMD, Training Device Design	Institutional, Skill	No	No

*Media, Methods, Delivery Systems

VARIABLES AND CONSTRUCTS OF PREDICTIVE AND PRESCRIPTIVE CTEA

As defined earlier in this report (Chapters 4 and 5), predictive models are concerned with the quantitative variables of a training program, TD/S or CBI system such as performance, training time, transfer, instructor time and the like. Prescriptive models and constructs emphasize selection of media, methods and delivery systems for entire training programs or for TD/S design.

Predictive CTEA

The variables listed below are those abstracted from model reviews as relevant to predictive CTEAs These variables need to be estimated in the conceptual phases of WSAP, ISD and TD/S and "fine tuned" as systems and training are developed and fielded.

All variables may not be relevant to a particular modeling effort but should be considered. Also, many of the predictive and prescriptive variables may be analyzed in conjunction with one another or, for purposes of a division of labor, may be estimated separately. It is assumed that predictive and prescriptive documents will come together at successive phases of the WSAP, ISD, and TD/S development processes.

The predictive variables are:

1. Performance effectiveness estimates for tasks, TD/S and/or the course as a whole; weights for combining task estimates or for combining various measures of the same task, such as:
   (a) knowledge training measures,
   (b) skill training measures

2. Time estimates for each task, TD/S and course as a whole differentiating TD/S and WS time, where appropriate

3. Student throughput:

   (a) Student input levels
   (b) Number of students entering training
   (c) Percent passing
   (d) Number of students completing training
   (e) Number and percent of students requiring remediation by task

4. Number of instructors and instructional support personnel

5. Number and practice time needed on WS required for training with allowance for down-time

6. Number and practice time needed on TD/S and or CBI stations required for training with allowance for down-time

7. WS vs. TD/S trade-off analysis or TD/S transfer of training estimates

8. Instructor/student ratios

9. Instructor/equipment ratios

10. Student/equipment ratios

11. Instructor/student/equipment mixes

A training resource management model may be appropriate to the management of mixes of student flows, instructional personnel, media, TD/S, CBI stations and WS. MODIA is the only existing model identified that has been used in this area. Kraft and Farr's (1984) model appears to be addressing similar concerns for CBI. TDDSS has identified bottlenecks and prescribed a family of TD/S as a solution.

Prescriptive Constructs

1. Descriptive and metric methods of selecting media, methods, and delivery systems based on task analysis, learning algorithms and media/methods lists

2. TD/S design models

3. CBI design models

4. Instructional management

APPLICATIONS AND TYPES OF TRAINING

As noted in Table 7-1, most CTEA models have been developed for institutional training in formal schools or for TD/S skill training. Further consideration needs to be given to more comprehensive training planning beyond the first formal course such as that done by TDDSS/TDDA. The potential advantages of early planning of all types of training are:

a. Relegation of tasks to an appropriate place in a comprehensive training plan.

b. Relegating general training to formal schools and specific training to OJT experience.

c. Planning for skill acquisition vs. integration, where skill integration may be better achieved through job experience, field exercises, and OJT.

d. Consideration to certification training, qualification training, and mobilization refresher training.

e. More efficient and cost-effective total training plans.

f. Improved allocation of TD/S, WS and CBI systems among institutional courses and unit training.

PERFORMANCE AND TIME ESTIMATES

TDDSS/TDDA is the only institutional model that attempts to estimate performance for the course and each task and to weight the effectiveness measures (see Chapter 4). CBP estimates performance for TD/S but does not necessarily weight the estimates. JSEP plans to estimate performance for CBI systems. The prescriptive models and HARDMAN TRRA assume stated performance objectives will be achieved if the course is properly specified. TDDSS/TDDA makes a unique contribution in that SMEs compare and iterate design characteristics in relation to performance and time. We found no instances where more than one measure of performance was used for a task.

Performance effectiveness measures should not be confused with graduates or percent passing. Although often used interchangeably, performance effectiveness refers to the measures of knowledge (e.g., percent correct) or skills (e.g., hits on target) by which passing scores are established.

Time estimates for institutional training have been made for tasks and the entire course by TDDSS and HARDMAN TRRA, but are used differently. TDDSS uses a method (see Chapter 4) to estimate whether the course might be too long or too short and to adjust the design accordingly. HARDMAN TRRA uses total course time as a means for costing trainee and instructor costs and perform tradeoffs with hardware, manpower and personnel characteristics. It is not known if TRRA also iterates time in relation to course design characteristics.

The value of time as a measure is not necessarily as straight forward as it may at first seem. In cost models that use trainee and instructor time as a basis for costing, total course length, measured in workdays for pay and allowance purposes, is likely to be a major cost item. A long course incurs greater costs than a short course. Workday time estimates are not, however, adequate for training purposes as learning is better estimated in hours and should take into account "off-duty" time in evenings and weekends.

The issue of time value is also brought out in relation to Orlansky and String's (1979) findings that CBI systems save an average of 30% of the time required for "conventional" instruction. This time savings would have value to the Army only under the following conditions:

1. The course itself can be shortened by one or more days freeing students to return to their duty posts, reducing the number of instructors, and achieving savings in related personnel pay and allowance costs.

2. The time saved can be used more productively for study on other tasks by students and instructors, without changing total course time and costs.

These conditions have not been explored in the literature.

Using predecessor or similar training programs to establish course and task time estimates should also be approached cautiously. Off-duty, weekend, and remedial time need to be clearly documented to make predecessor or similar training programs useful for time estimation purposes.

Theoretically, design characteristics could be iterated in relation to time and performance estimates using sensitivity analysis methods. Sensitivity analysis methods could also be used to study input and output characteristics and the relationship between performance and time. As far as we can tell, none of the models has attempted all of these analyses.

STUDENT THROUGHPUT

Included here are the following variables:

1. Student input levels
2. Number of students entering training
3. Percent passing
4. Number of students completing training
5. Number and percent of students requiring remediation by task

Student input levels (such as MOS, experience, ASVAB scores) and the number to be trained are included in the HARDMAN manpower, personnel, and training model. TRAINVICE and DEFT estimate the level of student knowledge and skill prior to TD/S training in analyzing the training problem for transfer. TDDSS analyzes reading levels, learning and performance difficulty, previously acquired skills, and changes in training characteristics. TDDSS also analyzes student-instructor-equipment flow to improve sequencing and suggest areas in which sequencing might be improved or a family of TD/S might be employed. JSEP proposes to use number of students to determine the best mix of students, courses, instructors and CBI systems.

TEEM includes a recycle rate as part of cost estimating and provides for fluctuation of input. We found no other references to estimating remediation requirements. Policy issues, special training needs or differences in the quality of performance outputs call for analyses by student input levels. For example, lowering ASVAB or other POI prerequisites may result in:

1. need for an alternative delivery system,

2. increased course length, or

3. increased remediation.

These alternatives might increase costs of training and development. Otherwise lowering prerequisites may show up in poor end-of-course or job performance.

Student output measures are the number of students who successfully complete the POI and the percentage of entering students who successfully complete the POI. In most non-military school environments, the "quality" of student output is also given (i.e., high pass, pass; grade; or rank in class). The quality of student output may now only be obtained on a sampling basis for research purposes.

There appears to be a need to predict and correct prospective performance deficiencies. Prospective performance deficiencies might be anticipated by including in the analysis student input levels, remediation requirements and performance estimates. The model could then be iterated to anticipate remediation to avoid performance deficiencies. TDDSS appears to come closest to addressing these problems.

INSTRUCTIONAL MANAGEMENT

Quantitative measures of instructional management are the number of instructors and instructional support (e.g. testing, laboratory assistant, etc.) personnel. Only TDDSS, HARDMAN, TEEM and JSEP consider instructional personnel. TEEM counts instructors in costing. As Hawley (Personal Communication, November 1985) indicates, to obtain an accurate estimate of instructional personnel requirements and instructor-student-equipment mixes, the course design should distinguish between classroom training, "hands-on" (TD/S and WS equipment) training, and CBI. The instructor/student ratio in classroom training may be quite "elastic" in relation to teaching effectiveness. That is, equal effectiveness may be achieved with one instructor to as few as 10 students or as many as 25 students. The limiting factor to the instructor/student ratio in classroom instruction is usually the size of the classroom facility. Instructor/student ratios for "hands-on" training, however, will be dictated by the role of the instructor on the TD/S or the WS. The ratio may be 1/1 for pilot training, 1/4 for a tank exercise, or 1/2 for the Unit Conduct of Fire Trainer (UCOFT).

CBI systems claim to free instructional management time as one of their advantages, allowing instructors to attend to other matters. However, contrary to some claims, CBI is not "instructor free". Instructional management for CBI must take into account instructor time required for supervision, individual assistance time required to make necessary courseware updates, and time to keep the system operational. Instructor time requirements will also depend on the design of the CBI delivery system (see Chapter 6). An estimate of the instructor/student ratio obtained by Kraft and Farr (1984) for CBI systems showed a range of 1/25 to 1/40. However, the types of CBI design and instructional management roles were not specified.

Qualitative considerations in instructional management emphasize other factors. TRADOC REG. 350-7 (1 April 1985, pp. 1-5) outlines minimum specialized training for staff and faculty personnel in " ... the following competency areas: (a) A systems approach to training. (b) Counseling techniques. (c) Training aids. (d) Class management. (e) Methods of instruction. (f) Administration. (g) Principles/theories of learning. (h) Evaluation/examination. (i) Communication. (j) Presentation of instruction."*

These topics underscore the potential differences in point of view between instructional personnel and training developers. School staff need to consider the many variables involved in the actual delivery of instruction as opposed to the development of the media, methods and delivery system. The quality of student input, student progress in training, equipment availability, equipment breakdowns, scheduling, remediation, make-up sessions, testing, progress assessment, adapting prescribed training techniques, adapting instruction to conflicts with other Army duties and the like are among the variables that have to be considered in the instructional management process.

*A number of additional topics that may be useful for staff and instructor training are as follows: 1. Individual and group differences-planning, presentation and assessment. 1.1 Planning for several levels of student background, preparation and progress. 1.2 Noting progress and problems during the presentation to overcome blocks and deficiencies in knowledge or skills. 1.3 Remediation needs, diagnosis, scheduling, presentation and management. 2. Contingency planning for "fail-safe" training delivery planning around equipment downtime, schedule conflicts, classes and individuals requiring more than scheduled time and practice to achieve proficiency.

For the most part, instructional management issues in the development of new training programs or for new WS have not been given adequate or consistent attention in the CTEA modeling literature. It seems to be assumed that instructional management will be a constant for all methods of instruction, an assumption that is not tenable. The issues that need to be addressed in prescriptive and predictive models are as follows:

1. Involve instructor/staff personnel in the planning of the new training program. In the TDDSS model, training developers first develop a set of "ideal" or "unconstrained" course alternatives. These alternatives are then revised with instructor staff personnel vis-a vis school constraints to achieve an "integrated" course.

2. Develop an instructor/staff training program or briefing for new training at the development or fielding stage. Only Vaughan's (1979) early applications of the Navy HARDMAN method consider this need explicitly. General instructor proficiency is assumed. Training is only concerned with the new weapon system training program. Many observers have noted that training effectiveness in general and transfer effectiveness of a TD/S depend a great deal on how effectively they are implemented.

As far as we can determine, consideration has not been given to instructional management variables for training outside of formal schools, i.e., OJT, field exercises, refresher training or reserve training.

Instructor turnover, scheduling and management of instructor personnel have not been considered by any of the existing models. These variables will affect the number, quality and effectiveness of instructors needed for the system.

In summary, more attention should be given to estimating instructional staffing requirements, to involving school and instructor personnel in course design and to training instructors for new WS training. Other models may be needed to address school management issues related to instructor flow and resource management. Attention to instructional management issues could be a sound point of departure for avoiding performance deficiencies and enhancing student motivation.

MOTIVATION

Very little direct reference was found to motivational variables in the reviews of models examined. As we were unable to secure some of the detailed modeling documents in time for this report, we cannot say whether motivation is treated, how it is defined or what importance it has in the CTEA models. For these reasons, this discussion is limited.

One prescriptive design model, TEEM, lists attitude learning as one type of learning for which media or methods are to be developed. No definition was found. To make a clear distinction, the course objectives would have to specify verbal or overt behavior to serve as an indicator of attitude learning as opposed to knowledge or skill behavior.

The literature of instructional media and methods has referred to the use of motion, effective graphics, audiovisual presentation and use of hardware as means of gaining attention and maintaining interest in the material being presented. The psychological literature focusing on active and frequent response, feedback, knowledge of results, and corrective action suggests that these variables, when properly embedded in training, instill confidence in the learner and build positive attitudes toward learning. It is not now possible to determine the validity of these claims for lack of convincing research and reviews of literature. However, if assumed to be true, it is theoretically possible for all prescriptive design models to develop a "motivation" or "interest" index for various delivery system configurations.

Wheaton and his colleagues (1976) pondered the motivation problem in terms of "consumer acceptance" of a training device in their empirical analysis of TRAINVICE models. They were unable to offer any suggestions as to how provisions could be made in the model for estimating "consumer acceptance."

Klein (1984) in using the CBP approach for training devices notes that SME instructors freely comment about the utility of proposed device design features. Perhaps early instructor involvement in design would lead to greater instructor acceptance.

Orlansky's and String's (1979) review of CBI showed that in the studies measuring attitudes about two-thirds of students and one-third of the instructors held favorable attitudes toward CBI training. The students' positive attitudes could be accounted for by CBI's use of the psychological principles mentioned above. The poor showing among instructors could be attributed to their reduced role in the teaching process, frustrations with breakdowns, lack of confidence in the method, perceived threat of replacement by a computer and many other perceptions as yet poorly understood.

TDDSS involves instructors in design, but no mention is made of training. If the assumption is accepted that the effectiveness of a training program or TD/S depends in part on the way in which it is used by school personnel, the discussion of instructor involvement in the design process and instructor training earlier in this chapter suggests that this approach may provide a lead to motivational concerns.

Effective sequencing, effective resource allocation, and reduction of maintainability and reliability problems of equipment may also affect motivation by reducing student and instructor frustrations attendant to untimely, unavailable or damaged equipment, training bottlenecks, or an illogical instructional sequence. TDDSS gives attention to all of these variables. Design models give attention only to sequencing.

Unanticipated performance deficits that may be due to high reading level requirements are given attention by TDDSS. As far as we could tell, only TEEM considers remedial cycles, and only on the cost side. None of the models adequately anticipates individual differences in student input levels and remediation needs. Failure to anticipate these problems could lead to frustration of students and instructors. Perhaps methods could be found to augment the prescriptive models to take account of these variables.

The stress of battle conditions has also been discussed as an area of motivation. Response to the differences between peacetime training and actual battle is mentioned in TDDSS in their reference to designing training for mobilization conditions. However, the meaning of this reference is unclear. The rationale underlying the sequence between institutional training and field exercises is, in part, that training should gradually simulate the reality and stress of battle conditions. How well this is done is unknown.

In summary, motivation is a subject that is ill defined in the literature and, as far as we know, in the CTEA models. A number of leads are suggested in this review to aid in focussing the literature, empirical research and training effectiveness models. It is suggested that with further development prescriptive models might attempt to take the following into account:

1. Attitude learning.
2. Motivational or interest index of media, methods and delivery system configurations and psychological principles of active and frequent student response, feedback, knowledge of results, and corrective action.
3. A sequencing, resource allocation and maintainability/reliability model to avoid the frustrations attendant to equipment breakdowns and unavailability, training bottlenecks and an illogical instructional sequence.
4. Early involvement of school personnel and instructors in design and instructor training to obviate problems attendant to effective

implementation of training at the school.

5. Individual differences in student input levels, reading levels and remediation requirements.

6. Realistic simulation of battle conditions in training and field exercises to take account of performance under stress.

RESPONSIVENESS OF MODELS TO WS DEVELOPMENT AND POLICY CHANGES

Two reasons for developing formal predictive and prescriptive CTEA training models are to make human performance requirements responsive to newly developing WS and to be able to respond to changes in proposed manpower, personnel, and training policy. So far, many of the issues attendant to WS design have been given attention, but no catalogue of policy changes has been developed that would aid in developing modeling requirements. A policy change catalogue appears to be needed for manpower, personnel and training.

The emphasis in the early phases of WSAP has been to integrate WS engineering with human performance capability models for human factors analysis, manpower, personnel, training, and TD/S design and prediction. These models are in various stages of conception or development. What is not clear at this time is the ability of the models to analyze the cost and effectiveness implications of policy changes under consideration or to respond effectively to enacted changes in policy.

At this time, only a few policy concerns are apparent they are presented along with their model implications as examples for further development.

Situation and Policy	Model Implications
1. Timely design and development of training and TD/S needed to field training in conjunction with fielding of a WS.	Existing and developing CTEA training and TD/S models are attempting to address this issue.
2. Corrective actions are required for Weapon Systems developed and fielded that are not responsive to human performance concerns resulting in "skill creep" and WS that strain the limits of human performance capability.	The TORSSA concept and developing human factors, manpower and personnel models are attempting to address these issues.
3. The Army needs to respond effectively to changes in the civilian economy. Changes in the economy often affect manpower, in turn affecting recruiting and retention of the most qualified and skilled personnel.	It is not clear how well manpower, personnel and training models address these issues. If student qualifications for training decline or if attempts are made to retain personnel by transfer to a similar MOS, training design and delivery adjustments are implied. Only TDDSS appears to address these implications.
4. The Army needs to be able to respond quickly in the event of a "small war" mobilization.	Only TDDSS considers training under assumptions of rapid mobilization but it is not clear in what ways. Trained manpower shortfalls could affect design, performance, lowered levels of student input, school capacity, training equipment requirements, WS requirements and training time. Further attention needs to be given to how CTEA models might more clearly respond to a mobilization scenario.
5. Army institutional training needs to be lengthened to: a. improve soldiers performance b. reduce field commanders responsibilities for unit training.	Only TDDSS considers performance course time, design and relegation of tasks between institutional training and OJT. The evidence is not clear on how responsive this model has been to this issue. MODIA has addressed training resource allocations, but not across institutional and unit training.

6. Army institutional training needs to be shortened and made more efficient with more emphasis placed on OJT and field exercises. (This policy is the inverse of #5). — Same as #5

We emphasize again that these are preliminary ideas subject to further development. If taken at face value, it is apparent that little is known at present about whether existing or developing models are addressing policy issues 2 through 6.

SUMMARY

A further review of the models and modeling issues was undertaken in relation to predictive variables and prescriptive constructs abstracted from the literature. The following observations appear to be warranted.

1. TDDSS provides a predictive and prescriptive model which, according to the available reports, appears to be the most comprehensive in concept. The model addresses the issues of training design for tasks or an entire course;TD/S design; trainability; performance estimating; time estimating; weighting performance estimates; iterating design in relation to time and performance estimates; planning for non-institutional training; responsiveness to various phases of WSAP development;variations in student input levels; instructor involvement in design; sequencing instruction to avoid bottlenecks; recommending "families" of TD/S; defining qualifications of SME s ; reliability through consensus of 3 or more SME s ; and provision of model output to aid in the development process. It may also be the model that may be most readily adapted to deal with policy changes and motivational issues.

   It should be noted that these are apparent advantages noted in the reports of the model developers (Dawdy and Hawley, 1982; Fredericksen and Hawley, 1981). The extent to which they successfully accomplish each of these objectives can only be evaluated as the model is used more extensively.

As noted in Chapter 4, the model has not been validated and to do so may prove difficult as it attempts to serve many purposes. It may be difficult to use for the same reasons. Only a comprehensive evaluation or validation plan could fully test all of its features.

TDDSS does not attempt to predict transfer of training. It has not been applied to CBI systems or non-systems exportable training and the value for TD/S design is unclear. Its cost model is also considered to be in need of improvement.

Overall, however, it appears to address almost all of the issues needed in the long run for a generic CTEA model. Further development and validation will provide greater insights into its usefulness.

2. HARDMAN TRRA is oriented toward hardware, manpower, personnel, and training tradeoffs and produces a "quasi-training" program useful largely in this context. The authors (HARDMAN, undated) acknowledge that their approach is oriented to the short range insofar as training is concerned. Its contribution may be most important for the earliest CTEA analysis in WSAP.

3. Among the transfer of training models, DEFT appears to be making important advances over TRAINVICE, in reorganizing variables for greater face validity, and in providing for three levels of analysis that are responsive to different levels of input information. Although a few validity studies have been done on TRAINVICE, further analysis is needed on DEFT to clearly establish its reliability, its validity compared to TRAINVICE, and its ability to discriminate between "good" and "poor" designs. Improvements in the transfer models need to take account of incremental transfer, estimate practice time and predict performance on a specific device. Concepts from CPB or TDDSS may be helpful in this regard.

   In the long run, a generic model will require integration of transfer of training measurement with training device design. The transfer models do not include cost analysis. The Transfer Effectiveness Ratio would be appropriate for cost purposes.

4. Among the purely prescriptive models TECEP, TDDA and TEEM are reportedly the most advanced in development and application. They offer a much more limited concept of training effectiveness than TDDSS confining themselves to media, method and delivery systems specification. As performance and time are not used for effectiveness measurement, designs are not iterated in relation to these measures. Further, they give little attention to student throughput and instructional management and for these reasons are likely to be less responsive to most training policy changes.

Reliability and validity studies have not been conducted with the exception of an evaluation of TECEP early in its development. The utility of these models for TD/S or CBI design is not clear. As these models were developed for similar purposes, comparative construct validity studies would be appropriate to test the models for use in designing an entire course, a family or TD/S, CBI design, and exportable package designs.

5. Kraft's and Farr's (1984) JSEP for CBI prediction may provide useful leads for examining mixes of students, equipment, and instructors and for non-system training. However, it is still in development and no data are available. Further model development is needed for CBI systems and non-system training (Chapter 6).

Chapter 8

METHODOLOGICAL ISSUES IN TRAINING EFFECTIVENESS ANALYSIS FOR CTEA

INTRODUCTION

This chapter discusses methodological issues and directions for improvement in training effectiveness analysis for CTEA. The topics discussed include the following:

1. Information and data sources
2. Reliability
3. Usage demands
4. Construct validity
5. Predictive validity
6. Tests and performance measures
7. Trade off methodologies
8. Decision making under uncertainty methods
9. Generic vs. specific models

INFORMATION AND DATA SOURCES

Information and data sources in the conceptual and development phases of WSAP include:

1. The threat scenario
2. WS functional description
3. The task analysis
4. Subject matter experts (SMEs)
5. Predecessor or similar training programs
6. Databases
7. Research literature

In later phases of WSAP empirical data is obtained. It should be noted that empirical data are fallible, particularly in early fielding of the WS. Hence, the process is one of successive approximation.

Given the threat scenario, the WS functional description and the task analysis in the WSAP conceptual phase, SMEs are required to make judgments related to training design, trainability, and training effectiveness. In only a few cases did reports indicate the qualification of the SMEs. Wheaton et al. (1976), Klein (1984) and Rose and Martin (1984) in their reports on TRAINVICE, CBP, and DEFT give a descriptive account of qualifications such as position, education, experience, areas of expertise and knowledge of other WS training programs. Dawdy and Hawley (1983) in their article on TDDSS define minimum qualifications for trainers to serve as SMEs. These studies presented data for three to nine SMEs. One study showed that less experienced training developers gave ratings that differed from their more experienced peers. We found no clear reference to the number of SMEs used in HARDMAN TRRA, TEEM, TECEP or TCA other than use of experts supervising less qualified personnel.

Descriptions of SME's qualifications are suggestive of the level of knowledge and expertise required to use a model. However, at present these descriptions are difficult to interpret. In most studies the number of SMEs is small, making judgment difficult about qualifications as they relate to reliability, validity and ease of use. A coding system for SME qualifications would enable analyses to be made across studies of various model applications and, when SME numbers are large, enable correlations to be made of qualifications, reliability, and validity. The coding scheme might include:

1. Role and position in the project: study director, researcher, training designer, training developer, trainer, lead trainer
2. Level of education
3. Military or civilian
4. Specific areas of expertise
5. Years of experience in each area of expertise

These are preliminary ideas that would be the subject of another study. Ease of use could be judged in terms of team effort requirements to set up the problem (task analysis, selection of variables, revision of learning algorithms) and to execute the model.

Comparison-based methodologies in which predecessor and/or similar training programs are used as points of reference for SMEs to formulate judgments and estimates have been used by HARDMAN (undated), Klein (1984) CBP, TECEP, and TEEM. These procedures require a search for training programs and judgments of their appropriateness vis-a-vis the new threat scenario. Of course, these methods are not appropriate in those cases where the threat and WS are so unique as to render predecessor or similar training programs inapplicable.

Hawley (Personal Communication, 1985) has cautioned against overemphasis on predecessor and similar training programs as

SMEs may lose sight of the threat scenario and predecessor or similar training programs may themselves yield less than optimal performance. This problem may be obviated by close attention to the threat scenario and task analysis vs. the predecessor training program.

The problems of using predecessor and similar training programs are also compounded by lack of accessible, accurate, and detailed historical records. Performance data are often based on SME's recall or estimates vs. actual historical records. Centralized recordkeeping of task performance, number of repeat trials, time to criterion and the like is not commonplace, rendering unlikely the possibility of assembling objective historical records for statistical analysis. Because off-duty study time is not normally reported, a biased time estimate for new training may result. Methods and media reported for the course may not represent those used in practice or found most useful by instructors.

These problems may be effectively addressed however, by an intensive analysis of the predecessor or similar training program. Reportedly (Mirabella, Orlansky, Personal Communications, 1985), several efforts are underway in the Armed Forces to address this problem and develop an historical database. If successful, these efforts would be helpful in developing baseline data for comparison-based analyses.

Figure 8-1 shows a three-dimensional conception of the relationship of information requirements, logical analyses, and

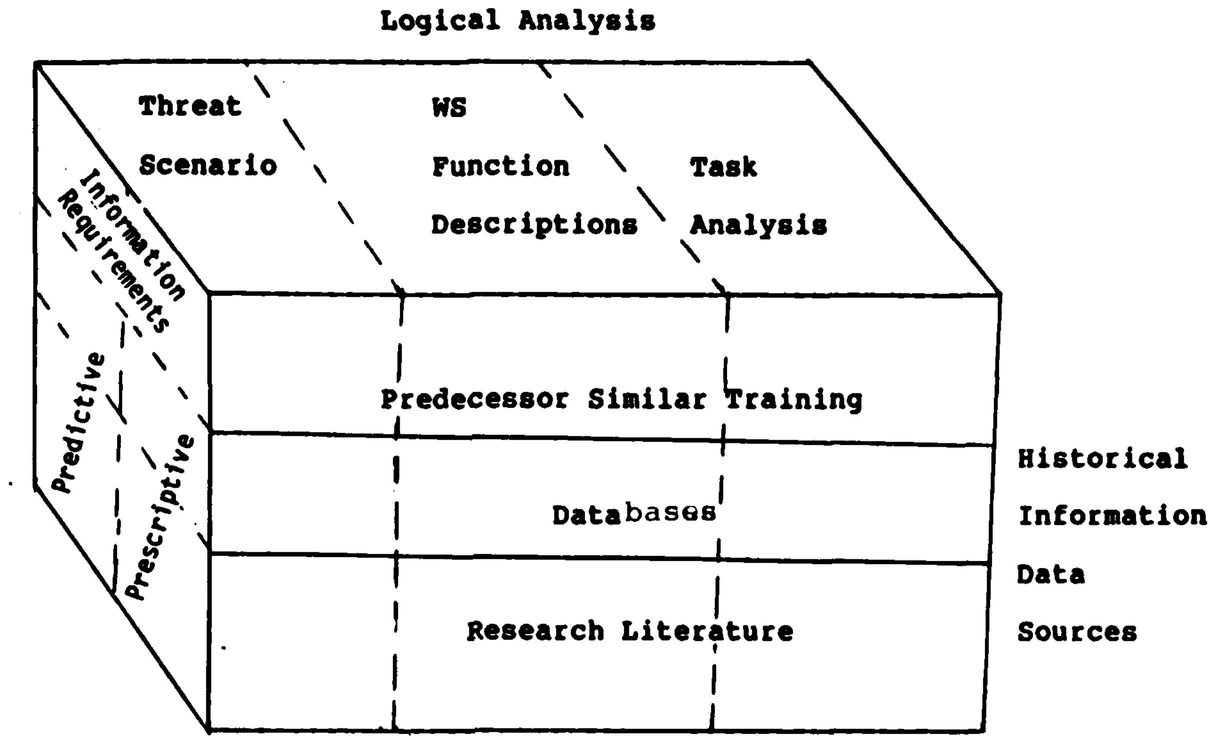

Figure 8-1. Three-Dimensional Conception of Information Requirements, Logical Analysis and Historical Information Data Sources.

historical information and data sources. Information requirements include the predictive and prescriptive variables and constructs discussed in Chapter 7. Logical analyses include the threat scenario, the WS functional description (or more advanced information, if available) and the task analysis. These logical analyses evolve in sequence. Historical information potentially includes predecessor or similar training programs, data bases or the research literature. The value of these historical information sources will depend on the similarity of the old and new threat scenario and WS. At the extremes, if the threat is substantially the same but a new WS is being developed to better meet it, the historical information sources are likely to be useful. However, if the threat and WS are entirely new, the historical information sources may be of relatively little value.

SMEs are not presented in the figure as it is SMEs of various qualifications throughout the Army that must make the required judgments An expanded version of this conception would overlay SME authority and qualifications as a fourth dimension.

From a training point of view, the threat scenario and WS functional description (or actual WS) are givens. In some cases, manpower and personnel information might also be available. A series of judgments can then be made of the relevance of the historical information sources to the threat scenario. Similarly, a series of judgments can be made of the relevance, reliability and validity of the historical information sources to the information requirements.

As the WS is fielded, new empirical data would gradually replace historical information sources and become part of the historical databases itself.

We emphasize that this is a conceptual framework as historical information sources and predictive and prescriptive models are not yet fully developed and organized. This conception, coupled with improved reliability and validity, should be helpful in guiding TEA study designs.

In summary, in practice, SMEs use their experiences, the threat scenario, WS functional description and task analyses, sometimes complemented by information about predecessor or similar training programs, to design training, assess trainability, predict effectiveness and estimate costs to meet the requirements of a proposed WS. A conceptual framework has been presented to relate information requirements, logical analysis and historical information.

RELIABILITY

Most models have not addressed the reliability issue; however, those that have done so report consensual agreement among three to nine SMEs (TDDSS, DEFT, TRAINVICE). With small samples of SMEs statistical methods of agreement coefficients are not useful. Reliability of all types of measures and constructs needs to be reported even if only in terms of the number of SMEs used, the degree of consensus, a minimal standard and methods of resolving differences among SMES. Reliability reporting is lacking in prescriptive methods where subjective judgments vis-a-vis media, methods and delivery systems may be critical, and in predictive methods where estimates of performance, time, students, instructors, TD/S and WS may rely on differing assumptions. Tests of reliability over time have not been conducted in any studies examined.

USAGE DEMANDS

The demands and difficulties of applying the models have only recently been given attention, and only then in anecdotal form. Klein (1984) describes the process of acquiring information for CBP for training devices and estimates the time required of an SME on a single device as about one hour but does not indicate time requirements to set up the problem. Faust et al. (1984) Wheaton et al. (1976) and Rose and Martin (1984) give the impression that TRAINVICE and DEFT may require one to three days of three to five SMEs for each device, that SMEs consider psychological ratings more difficult than other ratings; they consider the entire process demanding. Set-up time requirements were not indicated.

Hawley (Personal Communication, November 1985) considered TDDSS difficult to apply because of the wide range of coverage of the model. There is an implication in the HARDMAN TRRA documents (undated) that the "quasi-training plan" approach was devised as an "easy" method to apply in that what was needed was a short-term solution to CTEA to provide timely information for hardware, manpower, personnel and training tradeoffs. TECEP notes that lower echelon personnel can be effectively supervised to use the model. We found no other references to usage demands.

The models are in too early a stage of development to sacrifice usage demands to the more important concerns that give a model value - that is, timeliness, providing useful information in the WSAP, comprehensiveness in addressing important training variables and reliability and validity of estimates. As each of these goals are achieved, the model developers might also try to develop less demanding approaches. At present, more detailed documentation of staffing requirements, staff and SME training requirements, time (study directors, designers and SMEs) costs, and perceived difficulty of application will lead to greater clarity in usage demands and help pave the way for potential model improvements.

CONSTRUCT VALIDITY

Construct validity may be defined in a number of interrelated ways:

1. Operational validity: The extent to which the constructs or parameters of the model hypothesized to influence training effectiveness may be operationally defined from the research literature and training experience.

2. The comprehensiveness of the model in considering all identifiable constructs and parameters hypothesized to predict training effectiveness or prescribe training plans in various phases of WSAP.

3. Discriminant validity: The extent to which models discriminate between "good" and "poor" training plans or TD/S designs.

4. Comparative validity: The relative extent to which competing models effectively achieve their purposes.

5. The informational value of the quantitative and qualitative information yield for decisionmaking purposes in various phases of WSAP.

These definitions focus upon different meanings given to construct validity. The first two definitions reflect a theoretical and operational framework key to model construction itself and are often referred to as "face" validity. The last three definitions require empirical data about the models themselves.

Table 8-1 shows the different types of construct validity, the method of measurement (or development) and the models to which they apply. All models demonstrate operational or face validity. TDDSS is considered the most comprehensive in the range of issues it addresses, followed by HARDMAN with its emphasis on hardware, manpower, personnel and training tradeoffs. TRRA cannot be considered to be a comprehensive training model. The reader should note that comprehensiveness has not necessarily been demonstrated in application. Discriminant validity has not been established for any of the models. Knerr's (1984) review gives the results for a number of model applications. However, the results can only be interpreted subjectively. The absence of a sampling distribution, standard or tolerances for the metrics employed makes interpretation difficult. Part of the problem might be obviated if the reports gave results of the several iterations made. Comparison of these iterations presumably should show positive changes on the effectiveness metrics. Reportedly (Mirabella, Personal Communication, 1985), discriminant validity studies are in process for the transfer of training models.

Comparative validity studies for four TRAINVICE models and the need to compare DEFT and TRAINVICE were discussed in Chapter 4. The need for a comparison of the prescriptive models was discussed in Chapters 5 and 7 (summary).

Evidence related to the perceived value of information yield was not given in the reports we reviewed. The evaluations of HARDMAN and TECEP do not appear to be relevant.

A penetrating analysis of construct validity for the design models could be carried out by using an input, process, output analysis for each CTEA in the WSAP. Input to the model would include the detail and relevance of information of the threat scenario, the WS functional description and the task analysis. Process dimensions would be the media, method, stimulus and response variables incorporated in the models. Output dimensions would be the perceived information value of the training program descriptions and the metrics. An analytic effort could be used to narrow the comparison to two or three of the most promising models (TDDSS, TECEP, TEEM) for use in the Army. Empirical analysis could address the usefulness of the processes and outputs.

Table 8-1

Types of Construct Validity, Measurement and Applicable Models

Construct Validity	Measurement	Applicable Models
1. Operational Validity (face)	Operational definition from literature, training, experience, needs of decision makers in WSAP and training development.	All
2. Comprehensive Validity (face)	Above, plus comprehensiveness of variables considered.	Claimed but not fully established: TDDSS-Prescription and prediction. HARDMAN for hardware, manpower, personnel training tradeoffs.
3. Discriminant Validity	SMEs' ratings discriminate between alternative training designs, TD/S designs, and performance and time estimates for various designs.	Not established. Needed for all.
4. Comparative Validity	Relative extent to which models purporting to measure the same thing do so based on SME ratings and estimates.	TRAINVICE. Needed for TRAINVICE vs. DEFT; TDDSS/TDDA vs. TEEM vs. TECEP vs. TCA.
5. Information Validity	Perceived value of information by decisionmakers various phases of WSAP.	Not established. Needed for all.

The objective predictive measures (effectiveness estimates, time estimates, student throughput, number of instructors, numbers of equipment, and derivative measures) do not require further theoretical construct validation. It is the methods of estimation, (such as the use of SMEs of various qualifications or use of predecessor and similar training programs), weighting, combining, trading-off and reporting data that require further construct validation. In the literature available to us, only one method, Dawdy's and Hawley's MAUM (1983) was described in detail (see Chapter 4).

It should be noted that for the objective measures themselves, the method of agreement used for reliability is the same used for construct validity. The difference, if any, lies in the particular emphasis of a study. A single study, properly designed, could capture both reliability and construct validity.

PREDICTIVE VALIDITY

Predictive validity addresses the issue of the extent to which the model's output actually predicts the effectiveness of the training program.

Within the Army's Systems Approach to Training, a sequence of phased evaluations of training effectiveness provides the potential for empirical criteria for predictive validity studies. The types of early studies relevant to ISD have been variously termed developmental testing, pilot testing, verification/validation testing, and formative evaluation. These studies, generally performed on small samples, concentrate on areas of improvement needed in content, media, methods, delivery systems, instructional management, performance estimates and time requirements.

Later training evaluations are often conducted on larger samples in the first full-scale fielding of training and a new WS. Training evaluation after WS fielding is a continuing process directed toward combat readiness.

For purposes of validating SMEs predictive and prescriptive estimates, short range and mid-range criteria are the most practical. The further the criteria are removed from the predictors, the more likely intervening variables will confound the relationship. In ISD, continuing attention to training shortfalls after fielding must be distinguishable from the formative test.

Although many studies mention the provision of an audit trail, only those concerned with TD/S transfer have actually conducted predictive validity studies.

Comparisons of documentation from the conceptual phase to the development phase provide the first criteria. The closer the match, the greater the predictive validity. The development of certain data, however, may not proceed in a synchronous time line; for example, TD/S and CBI conceptual and developmental phases may well trail WS development and the fielding of the training program.

The predictive validity method requires follow-up data to relate "predictor" measures and "criterion" measures. Wheaton et al. (1976) cautioned that the predictor and empirical criterion measures must be compatible to enable interpretable analyses to be made. Ideally, both predictors and criteria should be formulated in advance for a sound study design, particularly to assure compatibility of measures and the availability of relevant criterion data. However, the follow-back designs and concurrent designs can probably be undertaken for most of the variables proposed here.

Table 8-2 shows criterion standards and actions in relation to the predictive variables and prescriptive constructs. A simple measure of absolute discrepancy between SME predictions and prescriptions (with or without predecessor or similar training programs) is the type of measure which is most practical to apply for objective measures when the number of SMEs in the predictors and the number of students in formative evaluations are small. When samples in the predictor and criterion variables are larger, standard errors and confidence intervals can be obtained to test for sampling error. All of the objective measures of effectiveness estimates, time estimates, student throughput measures, instructional personnel, and cost will lend themselves to this type of analysis. Tolerance standards for each type of measure should be developed to judge the importance of predictor-criterion discrepancies.

In early phases of developmental or formative testing the MAUM method might be useful to evaluate the importance of discrepancies and the need to improve the training program. When the training program is fielded, the number of students will be large enough to judge tolerance standards more objectively.

Table 8-2

Predictor-Criterion Analysis in CTEA

Predictive/Quantitative Variables	Criterion Standard and Actions
- Tests and Performance Estimates	Meets or exceeds standard Below standard-complete reevaluation
- Course and Task Time Estimates	Too long: shorten or provide additional practice Too short: lengthen POI, restructure tasks, search for more efficient training
- Student Throughput -- Input levels	Too low: provide more background, remedial or transition training Higher than expected-provide more depth and practice, or shorten training
-- Number of Students Entering, Number Requiring Remediation	Significantly higher than expected: Improve I/S ratio, consider tutoring, develop special support media
-- Percent Passing	If below standard-complete reevaluation
-- Number of Students Completing	If below manpower requirements, increase input, examine input levels as above; also see percent passing
-- Instructors and Support Personnel	If excess, reassign; if deficit, analyze capabilities, training, scheduling, assignments, ratios

Table 8-2 (continued)

Predictive/Quantitative Variables	Criterion Standards and Actions
- Number of Weapon Systems	Excess: reassign Shortage; examine downtime, scheduling, TD/S development and availability, I/S ratio, I/E ratio, S/E ratio
- Number of TD/S and CBI stations	Shortage: reexamine designs, distribution, scheduling, I/S, I/E, S/E ratios
- WS vs. TD/S Tradeoff; Transfer of Training	Increase or decrease practice time on each as appropriate
Prescriptive/Qualitative Constructs	
- Applications and Types of Training	Accurate identification; development and fielding in process or completed
- Media and Methods Selection	Degree of "match" from conception to development and fielding. Evaluate areas not matching in relation to predictive variables and redesign as appropriate
- Test and Performance Measurement Specification	
- Training Equipment System Configuration	
- TD/S Design; CBI Design	
- Distribution, Delivery, Instructional Management, Sequencing Analysis.	

The MAUM method should be designed to estimate the seriousness of discrepancies between predicted and early empirical estimates. The method can be derived from that used in TDDSS by Dawdy and Hawley (1982).

The importance of discrepancies beyond tolerance in effectiveness measures (for example, not to exceed 5% below standard in hits on target) are self evident: course time as a whole, number of students, number of instructors, and numbers of various types of equipment, impact effectiveness and design costs.

The qualitative constructs of prescriptive methods, that is, the descriptions of media, methods, sequence, types of training, and types of instructional personnel, can be compared using a matching process. The percent of agreement between the SME prescribed design and the developmentally tested or fielded product would be obtained. The lower the percentage agreement, the more changes in the transition from conceptual prescription to development to formative evaluation and fielding. A profile of the information sources employed and the variables and constructs would enable diagnostic analyses to be made of discrepancies that would feed into a databases that could aid future efforts in predictive and prescriptive CTEA.

TESTS AND PERFORMANCE MEASURES

The criterion-referenced testing model is appropriate to training program development. WS are conceived to meet perceived military threats or to gain a potential advantage in the event of an armed conflict. The threat scenarios are thus likely to represent situations and performance demands never before encountered but may have some similarity to previous threats. Hence, the performance criteria and the instruments needed to measure performance must also be developed to respond to the new threats as the WS and training are developed. The problem from a measurement standpoint has been referred to as criterion validity (Thorndike, 1971).

Criterion validity itself, particularly for new situations, starts with a logical analysis of "ultimate", "intermediate" and "immediate" performance requirements; the specification of conditions of each criterion; and the development of instruments, standards and tolerances to operationalize the criteria. This approach is not confined to the military; it is the universal first step in criterion validity development in industry and education as well. Empirical data do not exist directly on the new WS or its training program, so expert judgment bolstered by related research is required to define and operationalize the criterion measures. Thus, in the conceptual phase of WSAP the threat scenario, WS functional description, and front-end analysis lay the groundwork for specification of the criterion measures and instruments as well as the training program itself. Further, the criterion instrumentation, at least for formal training, is developed in conjunction with the training program, often by testing, measurement and training specialists with assistance from SMEs.

As a WS and its training program pass into the developmental and fielding phases, empirical data is obtained vis-a-vis the performance measures and training. This body of empirical data establishes a base for revising the performance measures or the training as the two must be in congruence with the training sequence and measures of the "ultimate" criteria of the threat scenario.

Predictive CTEA models, as far as we can tell, vary in the attention given to performance measurement specification. Vaughan's (1979) description of HARDMAN and Dawdy's and Hawley's (1983) description of TDDSS suggest that performance measurement specifications are included as part of the training program, but the level of detail of the specification is not clear from the articles available to us.

The prescriptive models that limit themselves to media, methods and delivery systems have not, as far as we can tell, given attention to performance measurement specifications.

Although the time required for testing and performance measurement may not always be clearly distinguished from training time, failure to provide for it could well lead to underestimates of course time requirements and result in other planning inadequacies.

A minimum for performance measurement specifications includes:

1. The measurable performance objectives for each task and sub-task as appropriate
   a. Situational givens
   b. Conditions and enabling objectives to be measured.

2. The types of performance measures to be used:
   a. Tests of knowledge and information
      (1) Objective, recognition or recall
      (2) Written
      (3) Oral
   b. Performance skill tests.
      (1) Observation checklist-mechanically enhanced
      (2) Performance recording devices

3. Scoring method

4. Standards for passing and remediation

5. Reliability and tolerances for certain measures.

By including the performance measurement specification, an analysis of congruence between the proposed training and the performance measures can be made and either or both revised if needed.

The construct validity of the performance measures in the conceptual phase of WSAP is reflected by their judged congruence with (a) the threat scenario (b) the task analysis and (c) the training design. The reliability of the skill measures (e.g., hits on target) can likely be estimated from the research literature and predecessor or similar training.

When "predictor-criterion" analyses to the development and fielding phases are undertaken, the "match" between the specifications and actual measures, actual training program and performance levels can be assessed. Coupling training design with performance measurement would then give clearer diagnosis of what needs revision- the training program or the performance measures.

TRADEOFF METHODOLOGIES

Tradeoff methodologies are apparent at a number of different levels:

1. Tradeoffs between WS design alternatives, human performance alternatives (human factors, manpower, personnel and training), and cost. TORSSA, HARDMAN and similar comprehensive models employ (or plan to employ) these methodologies to achieve systems with supportable costs and optimal performance potential.

2. Tradeoffs between cost and effectiveness in training design and prediction. Although of limited scope, almost all CTEA design models and prediction models analyzed a number of alternative training plans using "figures of merit" or performance and time measures in relation to costs. TRAINVICE and DEFT do not, however, have an associated cost model.

3. Design and effectiveness trade-offs were found in a limited number of training models: TDDSS in its applications to formal training vs. OJT, and consideration of design alternatives, and JSEP in its proposed analysis of CBI systems which is considering student flow, instructor/student ratios, instructor/equipment ratios, and multiple sites.

Notably absent were formal tradeoff analyses among time, effectiveness, and design characteristics. Much more could be done using tradeoff methods in this context.

DECISION MAKING UNDER UNCERTAINTY METHODS

Models focusing upon the conceptual and development phases of WSAP, by definition, are operating under conditions of uncertainty and risk as empirical data on the WS and training program under development will not begin to mature until prototypes are tested in the development and subsequent phases (Klein, 1984, HARDMAN). As far as we can tell, payoff tables and sensitivity analysis have only been used in cost-effectiveness analysis. Kraft and Farr (1984) propose to use sensitivity analysis and operations research techniques in their model, however, results are not yet available. Vaughan's 1979 description of the Training Requirements Analysis model used in an early version of HARDMAN briefly mentions use of a "fault-tree" analysis, a quantitative decision tree method used to detect potential faults in a training design. Unfortunately, details of the method were unavailable.

Conceptually, greater use could be made of sensitivity and trade-off methods in iterating training design with performance and time estimates; in determining the best mixes of students, instructors and equipment; and in non-system exportable training development.

GENERIC VS. SPECIFIC MODELS

The review of the CTEA literature presented in this report seems to indicate that at present all of the elements needed for a generic model are not yet in place. TDDSS presents a model much more comprehensive than the purely prescriptive models - TECEP, TEEM, or TCA- and presents major steps in the direction of a generic model. Their approach is flexible to adapt to various forms of task analyses, learning algorithms and training problems. It also uses some fixed elements. However, as noted in Chapter 7, TDDSS's applications have been limited and thus its claims to apply to a large variety of training problems are not fully validated. TDDSS does not attempt to measure transfer of training and its utility for TD/S design is unclear. The

TORSSA paper outlines a generic concept and points toward TDDSS and TDDA to implement it.

The prescriptive models are limited to training design and ignore performance and time estimates. They take a specific model approach rather than a generic approach.

Rosen, et. al. (1981) proposed a generic approach for Training Extension Courses (TEC) and non-system training. However, we were unable to find documents indicating that the approach was ever implemented. In 1980, Matlick, et. al. developed a Performance Guide for CTEA, however, it needs to be updated as a result of new model developments. Their approach directed model users to the potential usefulness of different models and presented guidance for adapting them to their needs. They note that the reliability and validity of the models was not known.

The adequacy of any of the models for TD/S is unclear. Besides TDDSS and the design models, a number of WS specific models have been developed (see Chapter 3 and Knerr) suggesting that an adaptable conceptual approach is sometimes needed as opposed to a fixed generic model that can serve all purposes.

Transfer of training models use structured formulas to predict transfer from a TD/S to a WS but have not yet been integrated with design models; have not as far as we know taken account of incremental transfer; families of TD/S and their sequencing; part-task trainers; instructor or student variance; or conditions in which the TD/S reasonably serves as the criterion.

Nonetheless, transfer of training is appealing in concept even beyond the confines of training devices. In Knerr's (1984) discussion, training transfer is sometimes used as the conceptual framework for all of training. Thus, it is possible to think in terms of transfer among sequences of enabling objectives; transfer from a formal course to OJT or unit field exercises (the job, according to Orlansky and String, 1979) or transfer in a sequence of courses.

The most promising leads for an integrated generic model for the moment appear to come from TDDSS and the transfer models. However, this integration is conditional on many of the issues raised throughout this report (information sources, reliability, validity, motivation, non-system training and the like) and the differences in the conceptual approaches used. TDDSS is both design and predictive and its formulae for effectiveness are based on MAUM weighted performance estimates and time estimates. Transfer models are predictive and use structured formulas based on transfer theory. The models could be merged if TDDSS can successfully demonstrate usefulness for TD/S and CBI designs, if validity can be more clearly demonstrated for each model, if the models can demonstrate their validity for non-system training, if they present information of value to decision makers, and if their metrics can be reconciled.

SUMMARY

This chapter reviewed a number of methodological issues in training effectiveness analysis in CTEA. Given proper attention, these issues will contribute to the development of more generic CTEA's.

BIBLIOGRAPHY

Adams, J. A. On the Evaluation of Training Devices, Human Factors, 1979, 21(6), 711-720.

Department of the Army, Armor Center, Program of Instruction, Basic Noncommissioned Officer Course, BNCOC 19K M1 Abrams. June, 1983.

------------ Armor Center, Combined Arms Training. April,1984.

------------ Armor Center, Combined Arms Training Operating and Support Costs. April,1984.

------------ Armor Enlisted Professional Development (Field Circular 21-309), July, 1984.

------------ TRADOC Pamphlet 11-8. Studies and Analysis Handbook, July, 1985.

------------ TRADOC Regulation 11-5. Army Programs, Cost Analysis Program (MOS Training Costs). November, 1977.

------------ TRADOC Regulation 11-8. Army Programs, Studies under AR 5-5. March, 1985.

------------ TRADOC Regulation 350-4. TRADOC Training Effectiveness Analysis (TEA)System. June, 1985.

------------ TRADOC Regulation 350-7. A Systems Approach to Training. April, 1985.

------------ TRADOC Regulation 600-4. Personnel-General, Integrated Personnel Support (IPS). June, 1978.

------------ TRADOC. One Station Unit Training (OSUT) Test Report. November, 1976.

------------ TRADOC Systems Analysis Activity,(TRASANA). Training Effectiveness Analysis (TEA) Handbook. 1980 draft.

------------ TRADOC Systems Analysis Activity, (TRASANA). Training Effectiveness Analysis: A Process in Evolution. (TR-49-82) October, 1982.

------------ Marginal Man and Military Service, A Review, Part 1. January,1965.

------------ Field Manual, Soldier's Manual, MOS 45K Tank Turret Repairman, Skill Level One/Two. July, 1976.

------------ Field Manual, Soldier's Manual, MOS 19K Skill Level Three, FM 17-19K3. September, 1983.

Awotua-Efebo, E. B. Guidelines for Cost-Effective Training Product Development. Educational Technology, March 1984,42-44.

Baker, F. B. Advances in Item Analysis. Review of Educational Research, 1977, 47(1), 151-178.

Balcom, J., Park, J., Toomer, L., of Dynamics Research Corporation and Feng, T. of the Army Research Institute for Behavioral and Social Sciences. Application of the HARDMAN Methodology to the Single-Channel Ground-Airborne Radio System (SINGCARS),Final Report, December, 1984.

Basic Armor Training Graduate Supervisor's Survey, Ft. Knox, undated.

Bastian, R., Edward, K., Medsker, K. & Schimmel, B. Formative Evaluation in Industry: Observations on a Course Development Project at Bell Laboratories. Performance and Instruction, pp.20-22, June, 1983.

Battalion Training Management System, Platoon Trainers Workshop. (BTMS-AC-83-1).

Berman, B. H. & Gorski, J. Producing High Tech Training: Turning SMEs into Trainers. Performance and Instruction, pp.11-12, April, 1984.

Bickley, W. R., & Bynum, J. A. Training Device Effectiveness: Formulation and Evaluation of a Methodology, U.S. Army Research Institute for the Behavioral and Social Sciences. September, 1980.

Blaiwes, A. S., Puig, J. A., & Regan, J. J. Transfer of Training and the Measurement of Training Effectiveness. Human Factors, 1973, 15(6) 523-533.

BNOC Organization on MICROTICCIT. Tests and Lessons. May 21, 1985.

Boldovici, J. A., Bessemer, D. W., & Haggard, D. F. Review of the M1 Unit-Conduct of Fire Trainer (UCOFT) Validation and Verification Test Report. U.S. Army Research Institute for the Behavioral and Social Sciences. June, 1985.

Braby, R., Micheli, G. S., Morris, C. L.,& Okraski, H. C. Staff Study on Cost and Training Effectiveness of Proposed Training Systems, Training Analysis and Evaluation Group (TAEG) Report 1. Naval Training Equipment Center. June, 1972.

Braby, R., Henry, J., Parrish, W., & Swope, W. A Technique For Choosing Cost-Effective Instructional Delivery Systems, Focus on the Trained Man, Final Report. Training Analysis and Evaluation Group (TAEG) Report 16. Naval Training Equipment Center. April, 1975.

Brett, B. E., Chapman, W. A., & Hawley, J. K. Cost and Information Effectiveness Analysis: Improved Methodology (Volume II) ~~Draft~~ Technical Report, Prepared for the Army Research Institute for the Behavioral and Social Sciences. April, 1982.

Brown, F. J., The Use of Simulation in Armor Unit Tactical Training/ Officer Tactical Training at the Armor School, Presentation at NATO Armor School Commander's Conference, France. September,1983.

------------ Total Force Tank Combat Training, presented in Boise,Idaho. August, 1983.

Casady, P. D.& Goodwin, G. J. The Implications of Computer Proliferation on Army Computer Training Strategies. Department of the Army. June, 1984.

Christ, R. E., Everett, J. E., Bryant, J. A., Acchione-Noel, S. C. & Caia, F. E. Lance Initial Screening Training Effectiveness Analysis, (TRASANA-TEA-3-83). Department of the Army, January, 1983.

Cicchinelli, L. F., Keller,R. A., & Harmon, K. R. Training Capabilities Test of Electronics Equipment Maintenance Trainer (EEMT) Findings and Conclusions,prepared for Naval Training Equipment Center Training Analysis and Evaluation Group by the Denver Research Institute. July, 1984.

Clark, R. C. Elam, R. J., & Merril, D. Training Content Experts to Design Instruction: An Adaptation of Component Display Theory. Performance and Instruction,September, 1983,10-15.

Czuchry, A. J., Doyle, K. M., Frueh, J. T., Baran, H. A. & Dieterly, D. L. Digital Avionics Information System (DAIS):Training Requirements Analysis Model, Users Guide, Volume II (TRAMOD). prepared for Air Force Human Resources Laboratory by Dynamics Research Corporation. September, 1978.

Dawdy, E. D.of TRASANA & Hawley, J. K. of Applied Science Associates. A Forecasting Method for Training Effectiveness Analysis. Proceedings of the Human Factors Society 26th Annual Meeting, 1982. 250-254.

Dawdy, E. E. & Hawley J. K. Trade-Off Resolution Support for System Acquisition (TORSSA): An Overview.(typescript) Army Research Institute Fort Bliss Texas Field Unit. December, 1984.

Decisions and Designs, Inc. MicroTICCIT Courseware for "Navigate from A to B": Instructor's Guide. May,1985.

Defense, Dept. of. Directory of Researchers for Human Research and Development Projects, April 1982.

Defense, Dept. of. Report on the OJT Study Task, On-The-Job Training in the Department of Defense, prepared under the auspices of the Defense Education and Training Executive Committee. January,1981.

Defense Science Board, Summer Study Briefing Report for Training and Training Technology. Office of the Undersecretary of Defense for Research and Engineering, July-August 1982.

Dray, S. M. & Eberts, R. E. A Guide to Evaluating Transfer of Training Studies of Army Training Systems. U. S. Army Research Institute for Behavioral Research and Social Sciences. January, 1982.

Dynamics Research Corporation. HARDMAN Methodology Handbook, Volumes I-IV,DRC Report Number E-5684. November, 1980.

------------ Application of the HARDMAN Methodology to the Enhanced Self-Propelled Artillery Weapon System (ESPAWS) Vol 1: Technical Report. Prepared for Army Research Institute for Behavioral and Social Sciences. April, 1981.

------------Vol II: Appendices.April, 1981.

------------HARDMAN, Comparability Analysis Methodology Guide, Volume III, Requirements Analysis, prepared for Army Research Institute for Behavioral and Social Sciences. Undated.

Eisele, C. R., Bell, T. R. & Laidlaw, C. D. Cost Analysis of Air Force On-The-Job Training: Development and Demonstration of A Methodology. CONSAD Research Company for Air Force Human Resources Laboratory. May, 1979.

Engstrom, G., Kestner, R., Laverne, J., Lineback, W., Saia, E. & Adame, M. TRASANA-TEA-1-84 Multiple Launch Rocket System Repairer, Cost and Training Effectiveness Analysis. Volume I, Main Report. TRADOC System Analysis Activity. February, 1984.

Faust, D. G., Swezey, R. W. & Unger, K. W. Field Application of Four Models Designed to Predict Transfer of Training Potential of Training Devices. U.S. Army Research Institute for Behavioral and Social Sciences. May, 1984.

Fitzpatrick, R. & Morrison, E. J. Performance and Product Evaluation. In R. L. Thorndike,Editor, Educational Measurement, Second Edition, American Council on Education, D.C., 1971.

Frankel, O. L. & Butler, R. A.The Army Manpower Cost System (AMCOS) Enlisted Resource and Budget Cost Models, prepared by The Assessment Group for the Army Research Institute for Behavioral and Social Sciences. May, 1985.

Frederickson, E. W . An Enriched User Oriented Cost and Training Effectiveness Analysis Methodology. The Proceedings of the Annual Conference of the Military Testing Association (23rd).Volume 1, 429-438. October, 1981. Applied Science Associates.

Gagne, R. M. Training Devices and Simulators: Some Research Issues. The American Psychologist, Vol.9, 1954. pp.95-107.

Gallegos, M. Method of Designing Instructional Alternatives (MODIA): Vol. 4. The Resource Utilization Model, R-1703-AF. A Project Air Force Report prepared by the RAND Corporation. July, 1977.

General Accounting Office. The Army Needs to Improve Individual Soldier Training in its Units, FPCD-81-29. March 31, 1981.

------------The Army Needs to Modify its System for Measuring Individual Soldier Proficiency, FPCD-82-28. March 30, 1982.

------------Army's Initiatives to Improve the Skill Qualification Test Program May Not Achieve Promised Changes, B-208585. Letter Report,August 18,1982.

Glaser, R. & Klaus, D. J. Proficiency Measurement Assessing Human Performance . In R. M. Gagne, Editor, Psychological Principles in System Development. Holt, Rinehart & Winston,N.Y., 1972.

Goclowski, J. C., King, G. F., Ronco, P. G., & Askren, W. B. Integration and Application of Human Resource Technologies in Weapon System Design: Coordination of Five Human Resource Technologies.(I) prepared by Dynamics Research Corporation for the Air Force Human Resources Laboratory, 1978.

------------ Integration and Application of Human Resource Technologies in Weapon System Design: Processes for the Coordinated Application of Five Human Resource Technologies.(II)prepared by Dynamics Research Corporation for the Air Force Human Resources Laboratory, March, 1978.

------------ Integration and Application of Human Resource Technologies in Weapon System Design: Consolidated Data Base Functional Specification.(III) prepared by Dynamics Research Corporation for the Air Force Human Resources Laboratory. May, 1978.

Goldfarb, R. S. & Mangum, S. L. Estimation of On-the-Job Training Costs for Satellite Communications Ground Station Equipment Repairers (MOS 26Y),Final Report for Army Research Institute for Behavioral and Social Sciences. September, 1983.

Hansen, D. H., Ross, S. & Harris, D. A. Flexilevel Adaptive Testing Paradigm: Validation in Technical Training, Prepared for the Air Force Human Resources Laboratory. July, 1977.

Hawley, J. K., Brett, B. E., & Chapman, W. A. Cost and Information Effectiveness Analysis: An Improved Methodology (Volume 1), Draft Technical Report prepared for the Army Research Institute for Behavioral and Social Sciences by Applied Science Associates. April, 1982.

Hawley, J. K., Frederickson, E. W. of Applied Science Associates. The Training Development Decision Support System: An Aid to the Design of Cost-Effective Training Programs. Proceedings of the Human Factors Society, 27th Annual Meeting, 1983, pp.397-401.

Hawley, J. K. & Thomason, S. C. Development of an Air Defense Cost and Training Effectiveness Analysis (CTEA) Methodology (for the AN/TSQ-73): Vol I - CTEA Within the Life Cycle System Management Model. U.S. Army Research Institute for the Behavioral and Social Sciences, December, 1978.

------------,Vol. II, Demonstration CTEA on the An/TSQ-73. December, 1978.

Heeringa, L., Baum, D. R., Holman, G. and Peio, K. An Analysis of Training Device Requirements Documentation Procedures. prepared for Army Research Institute for Behavioral and Social Sciences. October, 1982.

Hess, R. & Kantar, P. Method of Designing Instructional Alternatives, MODIA:Vol.5, A User's Guide to the Cost Model, a Project Air Force Report prepared by the RAND Corporation. October, 1977.

Hughes, R. G. Enabling Features Versus Instructional Features in Flying Training Simulation. Proceedings of the 11th NTEC/Industry Conference (Technical Report NAVTRAEQUIPCEN 1-M-306). 1978, pp.137-143.

Hughes, R. G., Mannan, S. T., and Jones, W. E., Application of Flight Simulator Record/Replay Feature (AFHRL-TR-79-52). Williams Air Force Base, Arizona, Air Force Human Resources Laboratory, Flying Training Division, 1979.

Johnson, W. B. & Fath, J. L., Implementation of a Mixed-Fidelity Approach to Maintenance Training, First Year Report, prepared for the Army Research Institute for Behavioral and Social Sciences, August, 1982-September, 1983.

Jorgensen, C. C. & Hoffer, P. L. Early Formulation of Training Programs for Cost Effectiveness Analysis, Army Research Institute unit at Fort Bliss, Texas. July, 1978.

Jorgensen, C. C. Early Training Assessment within Developing System Concepts, Research Report prepared for Army Research Institute for Behavioral and Social Sciences. August, 1979.

Kearsley, G. P., Seidel, R. J. & Wagner, H. Cost and Training Effectiveness Study of the Spatial Data Management System for Basic Education, Preliminary Draft, Prepared by the Human Resources Research Organization. September, 1982.

Klein Associates. Comparison-Based Prediction of Cost and Effectiveness of Training Devices: A Guidebook, Prepared for the Army Research Institute for Behavioral and Social Sciences, June, 1985.

Knapp, M. L. & Orlansky,J. A Cost Element Structure for Defense Training. Prepared by the Institute for Defense Analyses for the Office of the Under Secretary of Defense for Research and Engineering. November, 1983.

Knerr, C. M., Nadler, L. & Dowell, S. Training Transfer and Effectiveness Models, prepared by Human Resources Research Organization for the Army Research Institute for Behavioral and Social Sciences. January, 1984(a).

Knerr, C. M., Harris, J. H., O'Brien, B. K., Sticha, P. J. & Goldberg, S. L. Armor Procedural Skills: Learning and Retention. prepared by Decisions and Designs,Inc. February, 1984(b)

Kraft, R. H. of Florida State University & Farr, B. J. of Army Research Institute. Job Skills Education Program: Conduct Predictive Cost and Training Effectiveness Analysis, September,1983.

Kristiansen, D. M. A Job Aid for Modifying Ineffective or Inefficient Training Programs. U.S. Army Research Institute for Behavioral and Social Sciences. September, 1981.

Kristiansen, D. M. & Witmer, B. G. Guidelines for Conducting a Training Program Evaluation, (TPE). U.S. Army Research Institute for Behavioral and Social Sciences. September, 1981. (a)

Kristiansen, D. M. & Witmer, B. G. A Job Aid for the Systematic Evaluation of Lesson Plans. U.S. Army Research Institute for Behavioral and Social Sciences. September, 1981.(b)

Kruger, B. P. & Farr, B. J. Job Skills Education Program: Predictive Cost and Effectiveness Analysis. U. S. Army Research Institute for Behavioral and Social Sciences. undated.

Manning, D. T., Balsam, P., Ebner, D. C. & Brooks, F. R. Student Acceptance of Videodisk-Based Programs for Paramedical Training. Technological Horizons, November, 1983, 105-108.

Marcus, A. J. & Quester, A. O. Determinants of Labor Productivity in the Military, Center for Naval Analysis. Undated.

Matlick, R. K.,Berger, D. C., Knerr, C. M. & Chiorini, J. R. Cost and Training Effectiveness Analysis in the Army Life Cycle Systems Management Model, Prepared by Litton Mellonics,Inc. for the Army Research Institute for Behavioral and Social Sciences. September,1980(a)

Matlick, R. K., Rosen, M. H. & Berger, D. C. Cost and Training Effectiveness Analysis Performance Guide Prepared by Litton Mellonics,Inc. for the Army Research Institute for Behavioral and Social Sciences. September, 1980(b).

Miller, C. & Southard, L. D. Training Effectiveness Analysis -- A Process in Evolution. TRADOC Systems Analysis Agency. October, 1982.

NATO.Proceedings of the Symposium on the Military Value and Cost-Effectiveness of Training, NATO Conference, Defence Research Group, Brussels. January, 1985.

Navy, Department of. HARDMAN (Final Report), October, 1977.

Neal, G. L. Cost and Training Effectiveness Analysis. Educational Technology, XVIII, No.3,15-19, March, 1978.

Orlansky, J. & String, J. Cost Effectiveness of Computer-Based Instruction in Military Training. Final Report prepared by the Institute for Defense Analyses for the Office of Under-Secretary of Defense for Research and Engineering. April, 1979.

Orlansky, J. & String, J. Cost-Effectiveness of Maintenance Simulators for Military Training, prepared by the Institute for Defense Analyses for the Office of Under-Secretary of Defense for Research and Engineering. August, 1981.

Orlansky, J. The Cost Effectiveness of Military Training, presentation to the NATO Symposium on Military Value & Cost Effectiveness of Training. January, 1985.

Pieper, W. J., Guard, N. R., Michael, W. T. & Kordek, R.S. Training Developers Decision Aid for Optimizingh Performance-Based Training in Machine Ascendant MOS. prepared by Applied Science Associates,Inc., 1978.

Povenmire, H. K. & Roscoe, S. N. Incremental Transfer Effectiveness of A Ground Based General Aviation Trainer. Human Factors,534-542,15(b),1973.

Pressman, I. & Rosenbloom, B. CAI System Costs: Present and Future. Technological Horizons in Education, March, 1984, 94-98.

Ring, W. & Reif, H. B-1 Systems Approach to Training, Final Report. Appendix A: Cost Details, SAT-1,Vol.2, prepared by Calspan Corporation for the Air Force Human Resources Laboratory. July, 1975.

Rose, A. M., McLaughlin, D. H., Felker, D. B. & Magman, J. D. Retention of Soldiering Skills: A Review of Recent ARI Research. U.S. Army Institute for Behavioral Research and Social Sciences. March, 1981.

Rose, A. M. & Wheaton, G. R. Forecasting Device Effectiveness: I.Issues, Technical Report, Prepared by American Institutes for Research for the Army Research Institute for Behavioral and Social Sciences. December, 1984.

------------II. Procedures, December, 1984.

Rose, A. M. & Martin, A. M. Forecasting Device Effectiveness: III. Analytic Assessment of DEFT. Prepared by American Institutes for Research for the Army Research Institute for Behavioral and Social Sciences. December, 1984.

Rosen, M. H., Berger, D. C. & Matlick, R. K. A Review of Models for Cost and Training Effectiveness Analysis (CTEA). Prepared by Litton Mellonics,Inc. for the Army Research Institute for Behavioral Research and Social Sciences. March, 1981.

Rosen, T. Operational Development of Computerized Adaptive Testing. The Industrial Organizational Psychologist. Vol.22(3)1985.

Scribner, B. L., Smith, D. A., Baldwin, R. H. & Phillips, R. W. Are Smart Tankers Better Tankers: AFQT and Military Productivity, Office of Economic and Manpower Analysis, Department of Social Sciences, U.S. Military Academy, December, 1984.

Seidel, R. J. & Wagner, H. Cost-Effectiveness Specification for Computer Based Training Systems, Executive Summary, Prepared by the Human Research Resources Organization for the Defense Advanced Research Projects Agency, September, 1977.

------------ Ibid,Volume I, Development. September, 1977.

------------ Ibid,Volume II, Procurement. September, 1977.

------------ Ibid. Volume III,Operation and Maintenance. September, 1977.

------------ Front-End Analysis to Aid Emerging Training Systems. Human Resources Research Organization. February, 1980.

Shoemaker, H. A. Conditions Favorable to Successful Linkage Between Basic Research and Its Applications in Learning. Performance and Instruction, Dec. 1983-Jan.1984,3-6;9.

Smode, A. F. Training Device Design: Human Factors Requirements in the Technical Approach (Technical Report NAVTRAEQUIUPCEN 71-C-0013-1), 1973.

Softech,Inc., The Army Training and Evaluation System, Task 3 Report. Prepared for TRADOC and Defense Advance Research Projects Agency. June, 1978.

Solomon, H. Economic Issues in Cost-Effectiveness Analyses of Military Skill Training. Institute for Defense Analyses, IDA Paper P-1897, Task T-5-310. November, 1985.

Spangenberg, R. Planning, Developing and Validating Instruction, Chapter 5 ,Air Force Manual, undated.

Stephenson, R. W. & Burkett, J. R. On the Job Training in the Air Force: A Systems Analysis. American Institutes for Research prepared for Air Force Human Resources Laboratory. December,1975.

String, J. & Orlansky, J. Cost-Effectiveness of Flight Simulators for Military Training, Volume II:Estimating Costs of Training in Simulators and Aircraft, Prepared by the Institutes for Defense Analyses for the Office of Under-Secretary of Defense for Research and Engineering. August, 1977.

Temkin, S. Connolly, J. A., Marvin, M. D., Valdes, A. L. of Research for Better Schools, and Caviness, J. A. of the Army Research Institute for Behavioral and Social Sciences. A Cost Assessment of Army Training Alternatives. August, 1975.

Tufano, D. R. & Evans, R. A. The Prediction of Training Device Effectiveness: A Review of Army Models. Technical Report of the Army Research Institute for Behavioral and Social Sciences. April, 1982.

U. S. Congress, Congressional Budget Office, The Costs of Defense Manpower: Issues for 1977, Budget Issue Paper, January, 1977.

Vadhanapanich, S. Cost and Effectiveness Functions of Instructional Technology Systems: A Mathematical Model, a dissertation to the Florida State University, College of Education. March, 1976.

Vaughan, J. J. of Dynamics Research Corporation. Training Requirements Determination During Early Phases of Weapon System Acquisitions. 1st Interservice/Industry Training Equipment Conference Proceedings, pp.437-447.November,1979.

Vector Research,Inc. Cost and Training Effectiveness Analysis (CTEA): Handbook for Action Officers. BDM Services Company. January, 1976(a)

------------Development of Cost and Training Effectiveness Analytical Models/Methodologies for Assessing Training Development & Products. BDM Services Company, January, 1976(b).

------------General Model/Methodology for Conducting Cost and Training Effectiveness Analyses. BDM Services Company. December, 1975.

Wagner, H. & Seidel, R. J. Program Evaluation. Human Resources Research Organization. October, 1984.

Wheaton, G. R. & Mirabella, A. Effects of Task Index Variations on Training Effectiveness Criteria, Prepared by American Institutes for Research for the Naval Training Equipment Center. October, 1972.

Wheaton, G. R., Fingerman, P., Rose, A. & Leonard, R. Evaluation of the Effectiveness of Training Devices: Elaboration and Application of the Predictive Model, Research Memorandum 76-16. Prepared by the American Institutes for Research for the Army Research Institute for Behavioral and Social Sciences. July, 1976(a).

Wheaton, G., Rose, A., Fingerman, P., Leonard, R., & Boycan, G., Evaluation of the Effectiveness of Training Devices: Validation of the Predictive Model, Technical Report TR-76-12, prepared by the American Institutes for Research for the Army Research Institute for Behavioral and Social Sciences. October, 1976 (b).

Wheaton, G. R., Allen, T. W. & Johnson, E. of American Institutes for Research; and Drucker, E. H., Ford, P., & Campbell, R. C. of Human Research Resources Corporation; and Boycan, G. G. of The Army Research Institute for the Behavioral and Social Sciences. Evaluating Tank Platoon Battle Run Performance. May, 1980.

Wilbourn, J. M., Vitola, B. M. & Leisey, S. A. Trends in Training Performance: 1972-1974. Air Force Human Resources Laboratory. December, 1976.

Williges, B. H., Roscoe, S. N. & Williges, R. C. Synthetic Flight Training Revisited. Human Factors, 1973, 15(6), 543-560.

Witmer, B. G. A Job Aid for Structured Observation of Training. Army Research Institute for Behavioral and Social Sciences. September, 1981.

Wright, M. G. & McGarrahan, J. R. One Station Unit Training (OSUT) Graduate Evaluation for MOS 11D, 11E, 12B, 13B,16P and 36C. TRADOC Combined Arms Test Activity, Fort Hood, Texas, Final Report. August, 1976.

LIST OF ACRONYMS

ACA	Arayed criterion analysis
AE	Acquisition efficiency
ARI	Army research institute
ASSET	Acquisition of supportable systems evaluation technology
ATSC	Army training support center
BOT	Burst on target
CAT	Computerized adaptive testing
CBP	Comparison based prediction
CBI	Computer based instruction
CDB	Consolidated data base
CHRT	Coordinated human resources technology
CTEA	Cost and training effectiveness analysis
DEFT	Device effectiveness forecasting technique
DOD	Department of Defense
DRIMS	Diagnostic rifle markmanship simulator
EEA	Essential elements of analysis
ERIC	Educational resources information center
FSD	Functional system development
HARDMAN	Military manpower and hardware procurement
HMRT	Human factors, manpower, personnel and training
HUMMRO	Human resources research organization
ISD	Instructional systems development
JGD	Job guide development
LCSMM	Life cycle system management model
MAUM	Multi-attribute utility method
MMM	Maintenance manpower modeling
MODIA	Method of designinginstructional alternatives
MPMA	Mission resources management analysis
MTRR	Measures of training resource requirements
PLARS	Position location reference system
SME	Subject matter expert
SOC	System ownership costing
T	Transfer
TCA	Training consonance model
TDDA	Training development support system
TDDG	Training device desing guide
TDDSS	Training development decision support system
TD/S	Training device/simulator
TEA	Training effectiveness analysis
TEC	Training extension courses
TECEP	Training effectiveness and cost effectiveness prediction
TEEM	Training efficiency estimation model
TORSSA	Trade-off resolution support for system acquisition
TP	Training problem
TPA	Training program alternative
TRAINVICE	Methods for the analysis of training devices and simulators
TRAMOD	Training requirements analysis model
TRP	Transfer problem
TT	Training efficiency
TTFA	Training technology field activities
WSAP	Weapon system acquisition process