LEVEL II

ARI TECHNICAL REPORT
TR-78-A26

Improved Army Training and Evaluation Program (ARTEP) Methods for Unit Evaluation

VOLUME I: EXECUTIVE SUMMARY; STUDY DESIGN & FIELD RESEARCH

by

M. Dean Havron, Doris D. Albert &
Timothy J. McCullough

HUMAN SCIENCES RESEARCH, INC.
7710 Old Springhouse Road
McLean, Virginia 22101

and

Eugene Johnson III & Robert G. Wanschura
U.S. Army Research Institute for the
Behavioral and Social Sciences

NOVEMBER 1978

Contract DAHC 19-77-C-0001
Project: PR No. 76-68-PERI-OU
Engagement Simulation Technical Area
Prepared for

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

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

JOSEPH ZEIDNER
Technical Director

WILLIAM L. HAUSER
Colonel, US Army
Commander

Research accomplished
under contract to the Department of the Army

Human Sciences Research, Inc.

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-P, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR-78-A26 -VOL-1 (19) (18) ARI	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER (9)	
4. TITLE (and Subtitle): IMPROVED ARMY TRAINING AND EVALUATION PROGRAM (ARTEP) METHODS FOR UNIT EVALUATION. VOLUME I. EXECUTIVE SUMMARY; STUDY DESIGN & FIELD RESEARCH. (6)	5. TYPE OF REPORT & PERIOD COVERED: Technical Report, 8 Nov 76 - 31 Jan 78
	6. PERFORMING ORG. REPORT NUMBER: USR-RR-77/21-Ge -VOL-1
7. AUTHOR(s): M. Dean Havron, Doris D. Albert, Timothy J. McCullough, Eugene Johnson, III, and Robert G. Wanschura (ARI) (10) (14)	8. CONTRACT OR GRANT NUMBER(s): DAHC 19-77-C-0001 (15)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Human Sciences Research, Inc., 7710 Old Springhouse Road, McLean, Virginia 22101	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263743A773 (16)
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Training and Doctrine Command (TRADOC), Fort Monroe, VA 23651 (12) 62 p. (11)	12. REPORT DATE: November 1978
	13. NUMBER OF PAGES: 52
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for open release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

---

18. SUPPLEMENTARY NOTES

Research monitored technically by Eugene Johnson III and Robert G. Wanschura, Engagement Simulation Technical Area, Army Research Institute.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Evaluation
Performance
Army Training and Evaluation Program (ARTEP)
Tank/Mechanized Infantry Task Force

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

This study analyzes existing methods of implementing the Army Training and Evaluation Program (ARTEP) for a Tank/Mechanized Infantry Task Force. Principles of the ARTEP system need to be more fully developed before field units can use them to develop training and evaluation programs that effectively meet the goals of the ARTEP concept. Battalion field evaluation exercises should be planned with the ultimate goals of training diagnosis in mind. This means that significant emphasis must be placed on adequate training for evaluator/

Over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

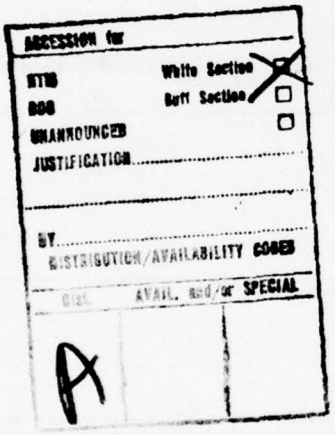
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

173250

Block 20 (Continued)

controllers. A three-volume report provides guidelines for meeting these requirements.

The report is intended primarily for readers interested in collective training, particularly those working with development of ARTEPs. Volume I is of general interest to ARTEP developers, training managers, policy makers, and users. Volume II provides data analysis and recommendations for refining current ARTEP implementation; it will interest developers, training managers, and policy makers. Volume III is a prototype guide for battalion-level use of the ARTEP. It is based on the recommendations from Volume II and is of interest to senior commanders, their staff, and officers who train others to perform evaluations and exercise control functions in the field.



FINAL REPORT
OVERVIEW OF CONTENTS

CONTENTS OF THIS VOLUME

EXECUTIVE SUMMARY

Problem Scope

This is a report of a study conducted by Human Sciences Research (HSR), Inc., for the Army Research Institute (ARI), entitled *Improved ARTEP Methods for Unit Evaluation*, under Contract DAHC 19-77-C-0001. The study focuses on one element of the Army Training and Evaluation Program (ARTEP)—the management of field evaluations of tactical units for purposes of training and diagnosis of training deficiencies. Attention is directed to the planning and conduct of field exercises according to Training and Evaluation (T&E) mission outlines provided in ARTEP 71-2, for training of Tank/Mechanized Infantry battalions.

ARTEP T&E Outlines; Research Requirements

The TRADOC-sponsored Army Training and Evaluation Program consists of concepts, guidance and training materials which together embrace all Army training. Guidance and training materials are still being produced, refined and dispatched to field users. Elements of concern to this study (mission T&E outlines) replace the Army Training Tests (ATTs) which, since World War II, provided the basic mission format for unit field training. The current ARTEP T&E outlines are performance oriented. They are designed to provide settings which more realistically represent realities of the modern battlefield, and changes in tactics required to accommodate new weaponry and supporting military hardware.

The transition to ARTEP is still in progress. Transitions from one system and set of concepts to another is never easy.[1] This study seeks to provide guidance and training materials that would assist in the transition, helping to assure that the concepts upon which ARTEP is based are wisely and imaginatively applied in unit field training.

Research Methodology; Problem Areas

The research approach selected is an iterative case study method. The cases consist of observations of the conduct of field evaluations from battalions of three divisions, and discussions with the key planner/supervisor of unit training in a fourth division. Information and data served to document current practices and user problem-solving efforts under the new ARTEP system. Comparison of data across cases permitted identification of recurrent problems.

Examples of problems and areas where guidance is needed are:

- Difficulties in interpreting and applying basic ARTEP system principles to unit training and evaluation programs.

---

[1] Major General John W. Seigle, "The Army Training System: A Status Report," DCST, HQ, TRADOC, AUSA Convention, October 1977.

- Uncertainties in developing the Evaluation Plan—particularly such aspects as exercise format, scenario construction, and technical simulation/control measures.

- Problems in interpreting and using the T&E outlines, and in organizing and training the Evaluator/Controller Group.

- Limited effectiveness in formulating, communicating, and using ARTEP evaluation results for remedial training.

Along with observation of cases, we interviewed cognizant officers at TRADOC/FORSCOM, schools and operational divisions. Two symposia on ARTEP were attended. Concepts and methods from the scientific literature (learning theory, systems analysis, psychometric methods, etc.) were reviewed to determine whether they might shed light on problems encountered, and for their contributions to solutions. Together, the scientific literature and military sources suggested directions for recommended solutions.

Problem Remedies

Remedial efforts took two forms. The first is a body of short-term guidance which is incorporated in the Field Guide. The second is identification of a range of issues wherein further ARTEP developmental research is still needed.

The Field Guide. Proposed solutions based on work described above are, of course, tentative. The Field Guide represents a first attempt to organize, and to incorporate proposed solutions into a practical document that officers who plan and implement battalion field evaluations will find useful. It includes:

- Procedures for the application of core ARTEP principles to decisions such as when to conduct external evaluations, how to task key personnel, how many assets to commit, and how to insure a "diagnostic" orientation.

- Guidelines for overcoming deficiencies in the most essential steps of preparing the Evaluation Plan and implementing it.

- An Evaluator/Controller Group module including a complete curriculum for training, and guidance for interpretation and use of the T&E outlines.

- Procedural improvements for the use of evaluation results.

Follow-on research and development. Follow-on research and development will undertake four tasks:

- Validate and refine the first generation Field Guide by field testing. This includes a number of subtasks. Among these are:

  – Development of instructions for evaluators as they rate performance on those items (standards) that require professional judgment.

  – Evaluation of revised formatting of standards, and their phrasing.

  – Evaluation of alternative means of combining item scores to validly reflect unit performance.

- Identify and evaluate alternative means by which battalion evaluations can be integrated into a program involving concurrent multi-echelon training.

- Determine how ratings obtained during battalion evaluations can best be utilized in remedial training.

- Update the analytic document (Volume II, this study) based on all work performed above.

Organization of this Report

This *Final Report* is presented in three volumes. These are:

Volume I – *Executive Summary; Study Design and Field Research*–describing project objectives, procedures and field data.

Volume II–*Analysis*–describing analytical steps to problem remedies.

Volume III–*The Field Guide for Evaluation Procedures*–presenting the first generation Field Guide in a three-module format.

PART ONE: STUDY DESIGN AND FIELD RESEARCH

CHAPTER I

INTRODUCTION

1-1. PROJECT OVERVIEW

This is a report of the first phase of a two-phase research project conducted by HSR for the Army Research Institute (ARI). Its objectives are to identify problems in the conduct of Army Training and Evaluation Programs (ARTEP) for field units, to develop remedies, and to incorporate these into a practical Field Guide to be used in conjunction with battalion field evaluations. Work is presented in three reports (Volumes I, II, III) of which this is the first.

Phase II research will extend the analysis of problems and recommended solutions, refine the Phase I Field Guide, and examine means of integrating battalion field exercises and resulting measures of performance into other training by operational units.

a. **Task organization.** The Phase I work performed is structured in three tasks, as follows:

(1) Task 1. Documents and collates current practices in ARTEP exercises. It accumulates the field experiences, the problems encountered and the solutions generated as a result of efforts to use the new ARTEP system and its external evaluation component by Tank/Mechanized Infantry units over the past several years.

(2) Task 2. Analyzes these data from the perspective of the problems they present for evaluating performance within the ARTEP framework. Task 2 also identifies possible solutions to these problems and recommends remedies.

(3) Task 3. Organizes these remedies into a Field Guide. The Field Guide is to be used by unit commanders, planners, trainers, and evaluators in conjunction with ARTEP 71-2 to better prepare and conduct evaluation exercises. Results desired are improvements in the conduct of training, and an enhanced ability to diagnose leader and unit deficiencies.

b. **Reports and summaries.** Project reports are presented in three volumes.

(1) Volume I. Executive Summary; Study Design and Field Research. This volume presents:

- Executive Summary.
- Study objectives, and research approach.
- Field research procedures used to collect information and establish a data base.

- Issues, problems and deficiencies in current practices of battalion field evaluations.

(2) Volume II. Analysis. Volume II describes the analysis performed in Task 2. This analysis is organized under three major headings.

- ARTEP system concepts and applications to Evaluation Plan design.
- Evaluation methodology.
- Use of evaluation results.

In each area, the analysis revealed that many problems involved in ARTEP implementation cannot be solved quickly or simply. As a result, each topical section separates these from problems for which immediate help could be found. Recommendations for how to deal with the more difficult problems are then presented. These recommendations and a summary of major results are reviewed in a concluding chapter of this volume.

(3) Volume III. Field Guide. The Field Guide is designed to address current problems described in Volume I. It incorporates proposed solutions to these problems developed in Volume II. It is a prescriptive document intended for unit planners, training managers and evaluators. It is organized so as to be used as a text by planners, instructors and by controller/evaluator trainees.

While prescriptive in tone, this is a first generation document. It would be unrealistic to believe, or to hope, that all solutions recommended are workable, or that all materials are precisely on target. In the Phase II follow-on study, it would be tested in the field. This test will undoubtedly lead to additions, modifications and deletions. These will be incorporated in a revised version.

1-2. PROJECT BACKGROUND AND MAJOR ISSUES

a. **ARTEP.** The Army Training and Evaluation Program (ARTEP) has been designated by the U.S. Army to replace the Army Training Test (ATT) and the Army Training Program (ATP) as the primary training instrument for battalion and subordinate level units. The purposes of ARTEPs are:

(1) To evaluate the ability of a tactical unit to perform specified missions under simulated combat conditions.

(2) To provide a guide for training by specifying mission standards or performance for combat-critical missions and tasks.

(3) To evaluate the effectiveness of past training of all echelons from crew/squad through battalion/task force.

(4) To provide an assessment of future training needs.

b. **ARTEP compared with ATTs.** Some comparisons between ATTs and the ARTEP T&E mission outlines which replaced them may be of interest.

(1) ATTs were training *tests;* ARTEPs, by contrast, emphasize training and training diagnosis of deficiencies.

(2) While the difference is one of degree, ATTs gave greater emphasis to mission planning, ARTEPs to performance.

(3) ATTs tended to be conducted as field drills, often well rehearsed in advance. ARTEPs, on the other hand, allow for logical branchings of activities within missions. Planners are encouraged to include surprises and occasions for innovative problem solving.

(4) It was felt that ATT checklists were over-developed and put too much emphasis on performance by lower echelons. Rating formats used by ARTEP are less well articulated, and place greater emphasis on command performance.

(5) Finally, ARTEPs are designed to reflect modifications in tactics that have occurred in the last quarter century.

ARTEP statements of performance criteria are a significant improvement over ATTs. However, it has not yet been demonstrated that ARTEP T&E outlines are stated with optimal completeness, specificity, and detail. TRADOC has been extending the guidance provided in the first five chapters of ARTEP volumes. However, to our knowledge, guidance has not been systematically checked for its value to planners. In ground tactics, the validity of evaluations must depend largely on professional judgments of evaluators. Thus, while the ARTEP approach is believed much superior to the ATTs, its objectives can still be compromised by poor field administration. The goals of this research are to help assure that ARTEPs, as practiced in the field, fully exploit their inherent training potential.

c. **Needs for research; focus of efforts.** Field observations and interviews conducted early in Task 1 served to help direct and sharpen the focus of research. These observations point to three needs which research must address.

(1) Unit needs for further guidance. Field experience with conducting formal ARTEP 71-2 evaluations quickly led to a number of problems of different types. The occurrence of these problems has often led to imaginative efforts by senior commanders, training managers, and evaluators—to deal with difficulties as they arose. Problems have also stimulated interest in being supplied sound guidance for improving ARTEP evaluation procedures.

(2) Need to broaden study scope. The central objective of this study, as stated initially, was to refine evaluation instruments (i.e., T&E mission outlines) and procedures for their use. Discussions with cognizant officers and field observations quickly led to a broader and far more diverse set of concerns. These were reviewed with the ARI Project Monitor, who accompanied the team on field visits and took part in interviews. It came to be realized that problems of improving the technical quality of instruments are deeply interwoven with many surrounding issues and problems. These had to do with organizing, tactically structuring, controlling and supporting the large-scale field exercises which serve as the evaluation vehicle. Problems also emerged in integration of ratings, and use of results to provide feedback. Thus, working solutions to (or at least a better understanding of) such issues were prerequisite to evolving solutions to a number of technical problems.

Accordingly, the scope of inquiry was broadened. Clearly, larger issues having to do with management of the evaluation vehicle, and interrelationships between them; and between management issues and technical problems needed to be better understood. Given this understanding, it should be possible to better understand the conditions under which evaluation instruments are to be applied.

(3) Need for practical guidance. The target audiences of the Field Guide are planners, training managers and evaluators. Guidance must recognize constraints that are virtually universal in the field. Among these are limitations in collective training time, personnel of the necessary ranks to serve as evaluators, logistics, pyrotechnics, etc., etc. Further, while scientific constructs may be found useful, their applications cannot presume education/training in areas from which they are drawn. Guidance must be set forth in terms familiar to users. The organization and the sequence of activities called for must be compatible with those to which users are accustomed.

1-3. SUMMARY

This chapter describes the study objectives, tasks and the organization of the report. Background information on ARTEPs was summarized. We have indicated how early field visits sharpened the focus of the study. The next chapter explains research methodology. Chapter 3 presents and discusses field observations. The analysis of these observations is presented in Volume II; the Field Guide is bound as Volume III.

CHAPTER 2

RESEARCH RATIONALE AND APPROACH

2-1. INTRODUCTION

This chapter describes the project objectives and shows how these objectives and "givens" required in the study gave rise to the research approach—an iterative case study method—and to research procedures. It then identifies sources of information and describes how information was integrated to produce the research products. The sources of information included observations of battalion field exercises, site visits, symposia, and military and scientific publications. These are documented in Attachment A.

The ARTEP system is in the early stages of its development. It was rapidly produced and dispatched promptly to field units. Some of the mission T&E outlines are still being refined. It has not been fully tested or evaluated. Decisions as to how best to implement the key concepts of decentralization, and of tailoring training to local mission needs often rest on rather broad guidance from TRADOC and proponent schools. Decentralization of authority has triggered a high degree of local initiative, as well as issues as to specific procedures for its implementation. It is well-recognized that substantial problems remain in implementation; indeed, it would be surprising if they did not.[1]

The objectives of this study stem from this context. An obvious requirement is to examine current practices in the implementation of ARTEPs in the field. It was necessary to determine what is being done by operational units to implement ARTEP, what types of problems they are encountering, and what directions might be pursued in problem resolution.

Since ARTEPs are very broad and all-encompassing, this initial effort was exploratory. Early inquiries suggested that no demonstrably valid data base was available, nor was it immediately apparent what such a data base should consist of.[2] The topical areas for data collection had to be identified, and the interrelationships between them determined. This required, in turn, a conceptual framework which would guide the collection of information and the assignment of items to defined categories. It was also necessary to develop concepts as to which categories of information were interrelated, and in what manner.

---

[1]Major General John W. Seigle, "The Army Training System: A Status Report," DCST, HQ, TRADOC, AUSA Convention, October 1977.

[2]Having completed the first year of study, we feel much better able to specify requirements for a data base.

This assessment gave rise to two concurrent efforts: (1) collection of information by observation of field exercises conducted by selected units, and (2) developing a conceptual model to structure the information and problem areas. Both efforts contributed to the end product—a Field Guide for the conduct of battalion-level external evaluations.

2-2. RESEARCH OBJECTIVES, REQUIREMENTS

The requirements of the study and the problems of field implementation of ARTEP help to further focus efforts:

a. **Selection of unit.** Project emphasis is on the operations of tank/mechanized infantry task forces. The lowest level at which armor, infantry and anti-armor are ordinarily brought together as coordinated maneuver elements is at battalion level. Field observations were then to concentrate on the broad level management of battalion field evaluation exercises, particularly:

(1) Policies bearing on the conduct of field evaluations.

(2) Evaluation planning.

(3) Field implementation of evaluations by evaluator/controller (E/C) groups.

(4) Use of results.

Thus, the tactical unit selected is the tank/mechanized infantry task force; the study focuses on operations of the evaluator/controller team in conducting field evaluations for this task force.

b. **Evaluation emphasis.** Consistent with stated ARTEP objectives, emphasis is placed on the use of field evaluations for training and diagnostic purposes *rather than* as measures of accountability or of operational readiness. (It appears likely that steps required to plan and conduct field evaluations are similar, whether the exercise purpose is training or to measure accountability. Uses to be made of results differ, however.)

c. **Requirements for model.** The generic requirements for the conceptual structure are as follows:

(1) The model should identify the key variables in the area of study. Items a., and b., above introduce a number of these variables. Other variables can be inferred from these.

(2) There should be well-defined categories to which variables are assigned. Rules for assignment of variables to categories should be explicit.

(3) Causal relationships are assumed among certain variables in the model. It should be possible to verify these assumed relationships. More specifically, within the context of management of field evaluations, it should be possible to relate variables within the areas of policies, planning, execution of field evaluations and use of results. If this can be done, we will have provided TRADOC/FORSCOM and division and brigade planners with the evaluation tools needed. These can be used to:

(a) diagnose deficiencies in current practices, and

(b) change policies and implement management procedures so as to improve the management of battalion-level field evaluations.

Note that all policies and measures taken focus finally on the operations of the Evaluation/Controller Group or team.

The variables and relationships emerged as field evaluation exercises were examined and the literature was reviewed. Here, the emerging model helped to ask questions, and to assign answers to variable categories. This process characterized the research approach and methodology described below.

2-3. RESEARCH APPROACH; METHODOLOGY

The research approach and methodology were derived from the study objectives and requirements just discussed. This is an iterative case study method. Cases consisted of observations of battalion field evaualtion exercises. On three occasions (Cases 1, 3, and 4) battalions operating as integral units were closely observed. Each consisted of observations of two battalions. To these were added reconstructions of methods by which other battalion field evaluations had been conducted from interviews with training managers (Case 2), and observations of company level field evaluations. A number of other sources, military and scientific, were woven around these cases. Figure 1 provides a schematic showing how all sources were drawn together to yield the research products.

The observations of battalions engaged in ARTEP external evaluations provided the basic data for this study. Observations started in December 1976 and continued until evaluations for six battalions and subordinate units had been observed. Data was collected on three additional battalions solely by interviews rather than observations. During these site visits, officers who planned and managed the evaluations were interviewed, and post-exercise critiques were attended. Where feasible, follow-up interviews were conducted with the managers and evaluators of the field exercises. The battalions observed were drawn from four separate divisions, two infantry and two armor. We thank these officers for their help and cooperation.

a. **Case observations.** Research procedures and lines of inquiry were developed as evaluations were observed. This is illustrated by recounting, *as examples*, a few themes as they emerged. Figure 2-1 shows how information from cases and other sources was combined. Divisions studied are designated as Divisions A, B, C and D.

(1) Case 1. Early observations of the first case—Division A—provided insights into field practices in planning for and conduct of evaluations. We had to assume these practices to be usual, since we had no firm basis for comparison at the time. For example, it became evident very early that use of evaluation instruments (T&E mission outlines) could be improved through better evaluator training. But it was also evident that this problem was interrelated with several others. One was that of controlling OPFOR so that all unit behaviors described in the T&E outlines would have occasion to occur. This exercise management problem appeared to merge with other problems of managing the field exercise.

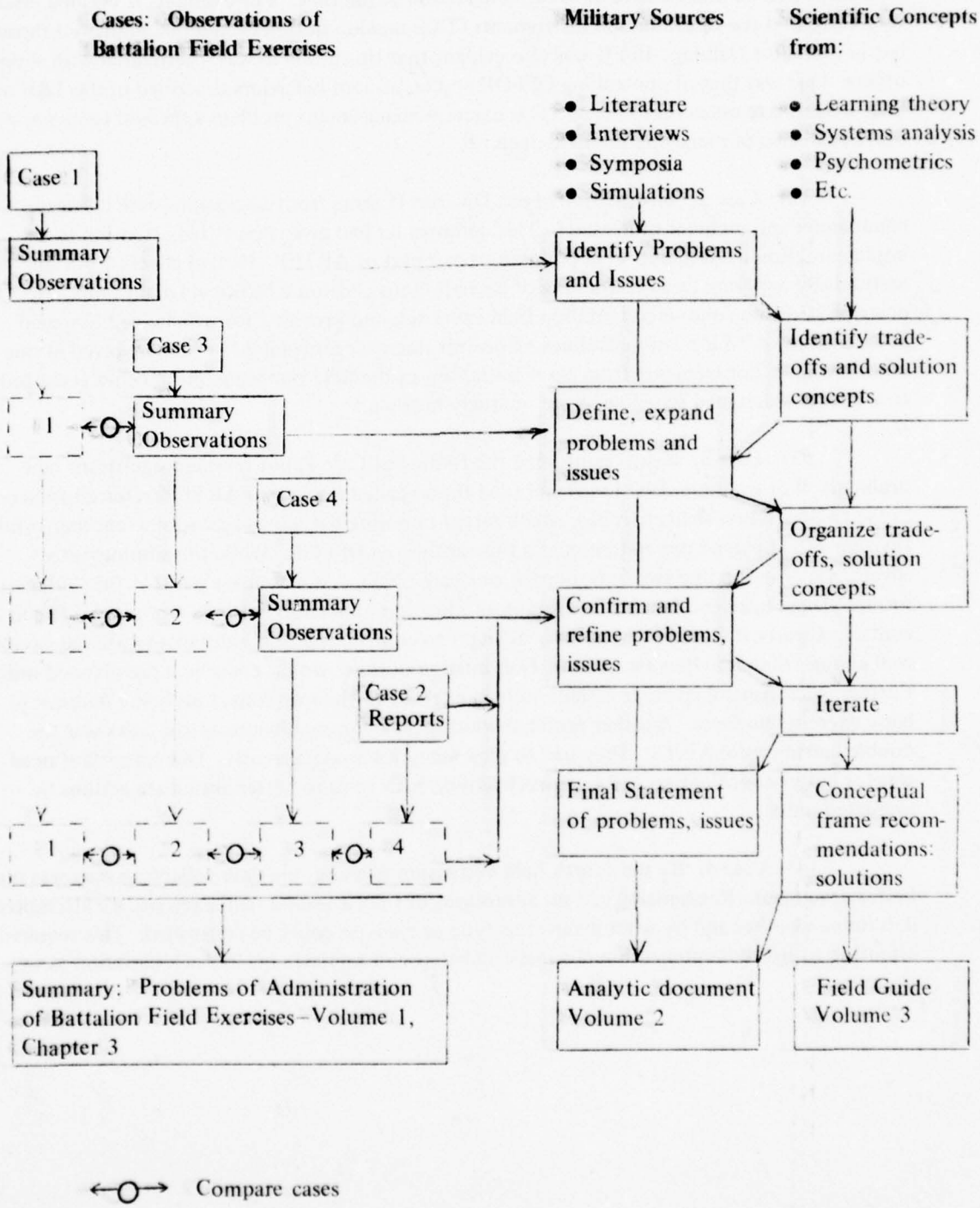
(2) Case 2. Information about Division D came from discussions with a brigade commander and member of his staff. This commander had given concerted attention to the implementation of battalion field evaluations as a part of ARTEP. He had checked out his methods by applying them to training of his battalions and some National Guard units. He described how he conducted battalion field exercises, and provided formats he had designed as extensions of T&E mission outlines. However, the procedures used—which involved at one point having major elements from *three* battalions in the field opposing one another according to a carefully designed schedule—were uniquely his own.

(3) Case 3. Case 3 reinforced the finding of Case 1, but revealed significant new problems. For example, Division B had used the so-called integrated ARTEP (referred to here as the two-battalion field exercise), which pitted opposing battalions against one another, while Division A had played one battalion at a time against an OPFOR. While the administrative advantages of evaluating two battalions at once are obvious, it was also clear that the "integrated" format placed heavier demands on evaluators and communication nets and was more difficult to control. Clearly it would be extremely difficult to control opposing sides in a two-sided exercise well enough that activities assumed in T&E mission outlines would occur in a pre-planned order. Further, such control could constrain units in exercise of their options, thus being inimical to basic exercise purposes. Another problem which became more obvious across cases was the double hatting among E/Cs. They had to play many roles concurrently. One important need was for better coordination and teamwork among E/Cs so as to better anticipate actions by evaluated units.

(4) Case 4. By the fourth field evaluation exercise, the data collection plan was much better developed. Recognizing certain advantages of the two-sided field exercise, we attempted to determine whether and by what means this type of exercise could be controlled. This required attention to E/C communication channels, and to use of tactical nets by each battalion as well.

FIGURE 2-1.

SCHEMATIC: RESEARCH METHODOLOGY



Procedures followed were similar to those used for Case 3. However, in addition, tape recorders were used to record communication traffic, including messages received in the TOC. (Sorting out recordings later and relating them to map positions of units on time lines proved to be an enormous problem, and one that was not adequately solved.)

b. Integration of information from cases. The above description of cases is intended to give some sense of the progression of the inquiry as we went from one case to the next. As such, it is not by any means an adequate description of all areas covered. These areas mentioned above are but a few of many.

Information from the four cases could now be compared. The Case 2 report is necessarily incomplete since it does not include field observations. These comparisons provide insights into how current operational practices are similar and how they differ, and problems common to all cases. Differences between cases permit definition of alternative command options. Similarities indicate problems that are more or less common across the board. Definitions of common problems helped to direct further inquiry and to guide searches for their solution.

Our data are presented and discussed in Chapter 3.

2-4. OTHER SOURCES OF INFORMATION

Two additional sources—military documents, cognizant officers and military training devices; and concepts from the scientific literature—contributed to the study.[3]

a. Military sources. Among military sources were "How to Fight" manuals, observations of the CATTs simulator and discussions with faculty members of C&GSC, and participation in two symposia on ARTEP. These helped to put field observations in perspective, to better appreciate constraints under which units must operate and to sharpen our definition of problems and issues. (We make this distinction between issues and problems: issues represent conflicts between desired goals. They require intelligent tradeoffs to be partially solvable. Problems are subsets of individual issues. Problems can be solved. The inherent conflict between emphasis on austerity and emphasis on combat realism is often an issue; effective training of evaluators is a problem.) Information from these sources was correlated with that drawn from case studies to help better define problems and issues.

b. Scientific concepts. The scientific literature provided concepts that seemingly might be applied to resolution of certain problems. (The proof remains to be established in Study Year 2!) As examples:

(1) Learning theory. Since the purpose of ARTEP missions is training and training diagnosis, concepts and principles from a well-established body of learning theory should be relevant.

---

[3]See Attachment A for references.

(2) Systems analysis. T&EO missions place three systems--evaluated unit, OPFOR and the E/C system--in dynamic interaction. With primary interest on the evaluator/controller group as a system, system analytic concepts should be applicable:

(a) as guides for forecasting likely system malfunctions, and for diagnosis of effectiveness of E/C operations, and

(b) in providing information about communication procedures needed within the E/C Group, and how this group can best monitor communications/actions of the battalion being evaluated.

(3) Psychometrics. Psychometrics can help to improve formats for T&EO standards and to improve the ways in which descriptive items might be better presented to evaluators.

Thus, military sources help to shed light on issues/problems identified in case studies. Scientific concepts and discussions with cognizant officers provide guidance as to possible tradeoffs that might be made to cope intelligently with issues. These sources serve also to direct the search for solutions to problems.

2-5. DESCRIPTIONS OF ISSUES AND PROBLEMS

As noted earlier, the model was developed to identify and classify variables important to the management of battalion field evaluation exercises. These variables and their interrelationships serve as background to a variety of issues and problems. Thus, variables and relationships are the "things" that problems and issues are all about. See the center column of Figure 2-1. It shows how issues and problems were identified, defined, refined and finally formulated by successive steps.

Issues and problems common across cases and which appeared most important were analyzed one at a time. For each issue/problem, information was organized as follows:

- Statement of the nature of the problem/issue.
- Citation of relevant documentary sources and data.
- Identification of key casual linkages to other issues in the matrix.
- Identification of major options for issue resolution.
- Statement of the key "pro's" and "con's" for each resolution option.
- Specification of the selected option and discussion of the supporting reasoning.

Attachment B presents an example of the logical format and contents of this analytical issue resolution process.

2-6. MATRIX FOR CLASSIFICATION OF ISSUES

All issues were classified into a single matrix shown as Figure 2-2. No single table or chart can convey the full complexity of this area nor reflect the many interrelationships between issues. There is also a rough antecedent-consequent relationship between entries in the four columns. The relationships between Columns 1, 2 and 3 are roughly correlated with roles and responsibilities of Army echelons in descending order.

While columns do not exactly map to formal functions of Army organizations, ARTEP system concepts which emanate from TRADOC/FORSCOM fall into the upper part of Column 1—the broadest category. These are the concepts and principles—the cornerstones of the ARTEPs—of which evaluation is one key element. The main tools available to TRADOC/FORSCOM for remedy of identified deficiencies are in formulating training policy guidance, and in guidance revisions. For TRADOC, such guidance may take a variety of forms including manuals, training circulars, etc.

A second class of issues (the lower elements of Column I and all of Column II) describes the basic options and limitations that senior commanders and training managers must confront in the course of planning and conducting external evaluations. This category includes issues such as deciding how many assets to spend, choosing the exercise format (e.g., single-battalion or integrated two-battalion approach), constructing the scenario, specifying the problem control procedures, planning the training of evaluator/controller teams, supervising their conduct of field exercises, etc.

Column III contains issues and technical problems of a more detailed nature. These issues/problems fall under the heading of formulating and applying the evaluation methodology. Among these are the appropriate task organization of Evaluator/Controller Groups, how Group functions are performed and integrated, evaluation criteria provided by mission outlines, procedures for their employment by field evaluators to rate performance, and so forth. Logically, midway between Columns III and IV, is a crucial problem area—namely how feedback from evaluator ratings can best be provided to members of evaluated units *during* field exercises.

The last issue category (Column IV) deals with formulating and communicating evaluation results, and the application of these results in revising training practices to correct unit deficiencies. Procedures in this area close the feedback loop which makes ARTEP a complete system for unit training and evaluation.

2-7. SUMMARY

This chapter summarizes the study objectives and research approach. An iterative case study methodology was used. Cases consist of observations of battalions conducting T&EO missions drawn from ARTEP 71-2. Military and scientific sources were used to help shed light on issues and problems identified in comparing data from the four cases. The more important issues/problems were analyzed one at a time. Military sources then provided directions and clues as to their resolution.

As an overview, issues are summarized in Figure 2-2, which contains four columns. The order of the first three columns corresponds roughly to responsibilities of TRADOC/ FORSCOM and field units. The fourth column reflects responsibilities of all levels of command in setting policy for, and providing information on performance to the units evaluated, thus closing the feedback loop. The matrix shown in Figure 2-2 is further articulated in the next chapter. Chapter 3 describes issues uncovered in field observations in some detail. Volume II suggests concepts and recommendations that may bear on their resolution.

Figure 2-2

CLASSIFICATION OF MAJOR ISSUES IN THE REFINEMENT OF TANK AND MECHANIZED INFANTRY TASK FORCE (ARTEP) EVALUATIONS

I. ARTEP SYSTEM CONCEPT ISSUES AND INITIAL APPLICATIONS	II. EVALUATION EXERCISE PREPARATION AND CONDUCT ISSUES	III. EVALUATION METHODOLOGICAL ISSUES	IV. RESULTS FORMULATION, COMMUNICATION AND USE ISSUES
A. Interpretation of Key ARTEP System Concepts 1. How operationally define "performance oriented" training? 2. How op. def. "mission/task oriented" training? 3. How op. def. "concurrent/multiechelon" training? 4. What are the appropriate forms and degrees of "decentralization"? (Also "realism," "austerity," "readiness," etc.). **B. Evaluation as ARTEP Tool** 1. How specify individual unit evaluation requirements? 2. How much relative emphasis and support for evaluations? 3. How to use eval. as training tool rather than proficiency test? 4. How balance evaluation and training goals during the field exercise? **C. Unit Contextual Issues** 1. How do TO&E and operational missions shape the evaluation process? 2. How do the nature and scope of other missions shape the evaluation process? 3. How do parent unit attributes shape the evaluation process? 4. How do local training facilities shape the evaluation process? 5. How do material asset constraints shape the evaluation process? 6. How do unit proficiency levels and ARTEP evaluation experience shape the process? **D. Evaluation Task Assignment Issues** 1. What are the appropriate degrees and modes of involvement of different command/staff echelons in the preparatory phase? 2. In the exercise conduct phase? 3. In the results utilization phase?	**A. Basic Field Exercise Format** 1. Which, how many units in what roles? 2. Where place evaluation priorities and emphasis? 3. If, when, and how to conduct "sub-unit" evaluations? **B. Scenario Construction** 1. Which, how many missions? 2. How to sequence and coordinate missions? 3. How shape scenario and missions overall for realism, doctrinal soundness and evaluation validity? **C. Tactical Problem Control** 1. What is the appropriate control element organization and procedure? 2. Engagement simulation: what basic approach to represent force ratios and weapons effects? 3. Engagement simulation: how employ specific simulation devices and techniques? 4. How play casualties and equipment losses? **D. Administrative Problem Control** 1. How to provide for range and safety requirements? 2. How incorporate on-line training activities, if any? 3. How provide for efficient administrative redeployment of evaluated units, etc.?	**A. Evaluation Group Attributes** 1. What size, composition and organization for each E/C Group? 2. How to select and train? 3. What other assets to provide? 4. How coordinate evaluation and control responsibilities? **B. Evaluation Criteria** 1. Reference T&E outlines; to what extent (a) explain or expand; (b) fill gaps; (c) guide rating aggregation? 2. What other criteria materials should be used, when, where? 3. How provide above materials on-line or elsewhere? **C. Observation Guidance and Procedure** 1. Where position to make observations? 2. By what means should observations be made? 3. What types of Group coordination and guidance are required during FEX? **D. Rating Procedures** 1. How distinguish and systematize observation/recording/rating behavior? 2. How coordinate inter-evaluator observation/recording/rating? 3. How guide procedures for overall unit mission and final formal evaluation ratings?	**A. Results Formulation** 1. What are the appropriate roles of each evaluation team member in the formulation process? 2. Should the same or "tailored" formulations be provided different users—units, command echelons, staffs? 3. How do the different chosen media affect content and formulation process? **B. Results Communication** 1. What are the appropriate audiences? 2. What is the optimal mix of feedback "loop sizes" for each audience? 3. What is the appropriate media mix for each audience? **C. Results Use** 1. What are the appropriate degrees and modes of involvement of different command/staff/unit echelons in interpreting results for training purposes? 2. For applying results to training schedules and procedures? 3. For conducting and evaluating revised training programs?

ATTACHMENT A

DATA SOURCES

Three primary sources of data were used to provide the information base for this project. They were:

(1) Field observations and interviews.

(2) Consultations and participation in workshops.

(3) Literature sources.

A-1. FIELD OBSERVATIONS AND INTERVIEWS

The major field research effort was the series of field observation and interview site visits undertaken since December, 1976. These efforts have resulted in direct observation of external evaluation exercises for seven battalions and surordinate units. Detailed reconstruction of procedures for three additional battalion evaluations was performed. Extensive interviewing of participants occupying all important roles in these exercises, and contact with many officers involved in evaluations for other battalions were also conducted. The battalions studied were drawn from four separate divisions–two infantry and two armor–representing a wide geographical dispersion within the continental United States. Future field work is tentatively planned for the observation of battalion evaluations in USAREUR, and for field assessment of the draft Field Guide in the projected second contract year.

A-2. CONSULTATIONS AND WORKSHOPS

In addition to field work directly related to the conduct of ARTEP evaluations, contact was made and coordinating meetings held with personnel from the following organizations:

- PM-TESS, Training Support Center, U.S. Army Traning and Doctrine Command, Fort Eustis, Virginia.
- Training Development Division, U.S. Army Training and Doctrine Command, Fort Monroe, Virginia.
- Historical Office, U.S. Army Training and Doctrine Command, Fort Monroe, Virginia.
- U.S. Army Combat Arms Training Board, Fort Benning, Georgia.
- Faculty, staff and students, Command and General Staff School, Fort Leavenworth, Kansas.
- Directorate of Training Developments, U.S. Army Armor School, Fort Knox, Kentucky.
- Deputy Commander and staff, the Infantry School, Fort Benning, Georgia.
- Cybernetics Technology Office, Advanced Research Projects Agency, DOD.
- Human Engineering Laboratory, Aberdeen Proving Ground, Maryland.
- Historical Evaluation and Research Organization, Dunn Loring, Virginia.

A-3. LITERATURE SOURCES

a. Training Doctrine and Guidelines

FM 21-6. *How to Prepare and Conduct Military Training*, Hq., DA, June 1974.

TC 21-5-1. *Training Management, An Overview*. Hq., DA.

TC 21-5-2. *Performance Oriented Training*. Hq., DA.

TC 21-5-7. *Training Management in Battalions*. Hq., DA, January 1977.

TC 71-5. *REALTRAIN*. Hq., DA, January 1975.

ARTEP 71-2 (Draft). *Army Training and Evaluation Program for Combined Arms Task Force*. September 1976.

ARTEP 71-2. *Army Training and Evaluation Program for Mechanized Infantry/Tank Task Force*. June 1977.

USACGSC. *Training Management Study Guide*. August 1977.

FORSCOM Regulation 350-1. *Active Component Training*. Current regulation, 1977.

FORSCOM Guidelines. *Training Guidelines FY 78*. May 1977.

TRADOC/FORSCOM. *ARTEP Workshop*. April 1976.

b. Tactical Doctrine

FM 71-1. *The Tank and Mechanized Infantry Company Team*. March 1976. ("How to Fight Manual")

FM 71-2. *The Tank and Mechanized Infantry Battalion Task Force*. July 1976. ("How to Fight Manual")

FM 100-5. *Operations*. July 1976.

FM 105-5. *Maneuver Control*. Hq., DA, December 1973.

USAITAD Report No. 14-0-76. *Military Operations of the Soviet Army*, United States Army Intelligence Threat Analysis Detachment, 25 May 1976.

TC 30-4. *The Motorized Rifle Regiment*. Hq., DA, June 1975.

TC 30-102. *The Motorized Rifle Company*. Hq., DA, September 1975.

c. Scientific Literature

Daniel R. Ilgen, *et al. Performance Feedback: A Review of its Psychological and Behavioral Effects.* (Army Research Institute, Technical Report No. 1., February 1977.)

M. Dean Havron. *Evaluating Combat Systems: Establishment of Criteria and Their Use in Selection of Key Systems Factors.* (Arlington, Va.: Human Sciences Research, Inc., March 1961.)

Charles Hitch. "On the Choice of Objectives in Systems Studies," in D.P. Eckman (ed.), *Systems: Research and Design.* (New York: J. Wyley, 1961.)

Clark L. Hull. *Principles of Behavior.* (New York: Appleton-Century-Crofts, 1951.)

Martin I. Kurke. "Operational Sequence Diagrams in System Design—Description of the OSD," in *Human Factors, 3,* No. 1, March 1961.

David Magnusson. *Test Theory.* (Reading, Mass.: Addison-Wesley, 1966.)

Stanford L. Optner (ed.). *Systems Analysis.* (Baltimore: Penguin, 1973.)

TRADOC Pamphlet 350-30. *Interservice Procedures for Instructional Systems Development.* (Author, August 1975.)

Harold P. VanCott and James W. Altman. *Procedures for Including Human Engineering Factors in the Development of Weapons Systems.* (Wright Air Force Development Center, October 1956.)

ATTACHMENT B
EXAMPLE OF ISSUE RESOLUTION LOGIC FOR EVALUATION PLAN DESIGN

Issue*	Related Issues	Issue Resolution Options	Option Pros and Cons	Issue Resolution
II.A.1. Which, how many units in what roles? (This issue has two aspects. Addressed here is that involving the choice of a one battalion or two battalion "integrated" format. The other aspect, dealing with the specific composition of each Bn task force, is not included in this example.)	**I.A.4**: ARTEP System Concepts directly applicable include: (a) **decentralization**, since the two Bn format may make Div. sponsorship preferable. (b) **realism**, since each alternative format has pros and cons re: fidelity of tactical environment. (c) **austerity**, since the relative cost of either format is a key contingency. **I.B.3**: Which format is more conducive to the diagnostic purpose of the evaluation? **I.D**: What do the two formats imply re: allocation of basic responsibilities in each evaluation phase? **II.B**: What scenario attributes are sensitive to these format alternatives? **II.C**: What exercise control functions are sensitive to these format alternatives? **III.A**: What E/C Group attributes are sensitive to these format alternatives? **III.C**: What E/C Group procedural guidelines are sensitive to these format alternatives?	***Option No. 1*** Non-integrated, i.e., one battalion evaluation exercise format. **Option No. 2** Integrated, i.e., two battalion evaluation exercise format. **Option No. 3** Integrated evaluation exercise format with *three (or more)* battalions concurrently in "round-robin" use of lanes and OPFOR roles.	***Option No. 1*** **Pro**: Can be readily staged by Bde, promoting greater decentralization (I.A.4 and I.D). Does not require US Bns in opposition, promoting greater realism (I.A.4). Apparently less net asset requirements, promoting greater austerity (I.A.4). Apparently reduces competitive atmosphere/pressure, fostering diagnostic orientation (I.B.3). Requires less complex coordination of scenario (II.B). Requires less complex control capabilities (II.C). Requires smaller, simpler E/C Group structure (III.A). Permits simpler E/C Group procedures & greater emphasis of evaluation duties (III.C). **Con**: Requires dedicated OPFOR, with limited capability to simulate correct force ratios (I.A.4). If Bde sponsored, increases coordination requirements with Div. on scheduling, support, Plan approval (I.D). **Option No. 2** **Pro**: If Div. sponsored, gives Plan drafting role to more experienced staff (I.D). If Div. sponsored, may save staff time by drafting one basic Plan for all Div. Bns. (I.A.4). **Con**: Converse of all positive attributes cited for Option No. 1. **Option No. 3** **Pro**: Same as Option No. 2. **Con**: Same as Option No. 2, but generally on an intensified basis.	***Option No. 1***

*See Figure 2-2, Page 15, for key to index numbers used in this example.

CHAPTER 3

SUMMARY OF CASES: FIELD OBSERVATION DATA

3-1. CURRENT ARTEP FIELD EVALUATION PRACTICES

a. **Introduction.** Field observations during the first year of this project produced data from six battalion-size field evaluation exercises. This section discusses the broad trends identified from this data. Detailed documentation of ARTEP deficiencies is provided in Section 3-2.

b. **ARTEP System Concepts.** The implementation of ARTEP as *a complete system* for unit training and evaluation requires that field units identify the basic principles of ARTEP *as a system* and that they apply these correctly. Failure to do so has three possible consequences: (1) ARTEP is viewed as being indistinguishable from ATTs; (2) ARTEP's unique features are overlooked and thus not used; and (3) ARTEPs are used inconsistently across units in developing training policy. Such consequences were observed!

c. **Evaluation Exercise Preparation and Conduct Issues.** Current field practices in the preparation and conduct of evaluation exercises indicated a lack of awareness of the tasks' complexity, or the need to provide for the increased complexity. For example, some units continue to use the two-battalion opposing force's format although the one-battalion format is less complicated and therefore more easily managed. Most field units attempted to run an excessive number of missions in too short a period of time, which resulted in numerous E/C personnel interventions. Both of these deficiencies increased the need for better control during the exercise. Yet adequate control procedures were not provided for in the Evaluation Plan.

d. **Evaluation Method.** Most evaluator teams consisted of 25 to 30 personnel with no distinction being made between evaluation, control or simulation functions. These personnel were typically chosen from the internal assets of the parent brigade, although personnel were also drawn from elsewhere in the division. In general, training of the evaluator team was perfunctory and never exceeded a single day.

e. **Use of Results.** Most of the units observed made no formal provision for critiques by individual evaluators of the units to which they were assigned, either on-line or immediately following the exercise. Common practice was to hold a large coordinating meeting after the exercise, in which ratings were finalized and transmitted in summary form to the sponsoring commander.

3-2. DOCUMENTATION AND DELINEATION OF KEY AREAS OF DEFICIENCY

Table 3-1 summarizes results of the observations of selected field unit ARTEP evaluations. This table is keyed to the scheme of specific issues and types of issues introduced in the issue classification matrix presented as Figure 2-2 in the previous chapter. The data elements in individual cells correspond to specific observations for specific units.[1] These focus on unit activities which were seen as "problems" or deficiencies in the use of ARTEP principles and official guidelines. The following paragraphs detail these observations, and cite common patterns and tendencies. The construction of an overall framework for the organization of the data *followed* these tentative efforts to find commonalities in the data collected from the field.

a. ARTEP System Concept Issues and Initial Applicatons

(1) Interpretation of Key ARTEP Systems Concepts. Under this heading, four areas of weakness were identified during observations of ARTEP evaluation exercises. These areas were:

(a) Confusion about tailoring the ARTEP to fit individual unit needs.

(b) Lack of understanding of techniques for application of specific T&E standards for evaluation and training.

(c) Tendencies toward "cycling" and "peaking" in the overall training/evaluation schedule, despite ARTEP emphasis to the contrary.

(d) Uncertainties about the appropriate nature and degree of responsibilities of specific HQ/staff echelons in the overall effort and specific phases.

There was a tendency by most performing units to treat ARTEP 71-2 as the last word, rather than as a broad advisory outline and flexible tool. This resulted in an often unimaginative, literal, and overly mechanical application of the T&E outlines to the unit's training/evaluation program as shown in Table 3-1, under I.A.1.a. As a result, efforts to tailor ARTEPs to specific unit training weaknesses, missions, and local circumstances appeared sparse. It was also evident that this was reinforced by a lack of definitive Army-wide guidance for conduct of field exercises and application of specific T&E standards to unit training or evaluation. Lack of guidance led to overdependence on individual evaluators' interpretations and judgments during the formal evaluations. This problem existed among most performing units, as shown

[1] The categories shown in the table evolved as we observed successive cases. We had but few of them as reminders when we made field observations. Many entries into cells are based on the best collective recollections of observers from the research team. Finally, entries in the table, and in the accompanying explanatory discussions cite what we conceive as *deficiencies* in current practices. These deficiencies are the focus of the study. Effective practices receive less attention.

Table 3-1*
TABULAR SUMMARY OF BN ARTEP EVALUATION EXERCISE DATA ELEMENTS

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
I.A. Interpretation of Key ARTEP Systems Concepts	1. Questions about how much "tailoring" of ARTEP 71-2 materials is required by field units.	a. Over-reliance on literal (mechanical) use of T&EOs.	a. Same	a. Same	a. Not applicable. (Extensive adaptation of materials to own requirements)
	2. Questions about how to apply/define specific T/E standards for eval. or training.	a. Overdependence on individual controller/evaluator interpretation (little concerted guidance).	a. Same	a. Same	a. Highly specific elaboration of specific Standard items.
	3. Cycling/peaking tendencies *in* overall T/E program.	a. Lack of *sustained* Bn level training prior to period of eval.	a. Same	a. Same	a. Not applicable. (Bn exercises conducted concurrently with subunit training).
	4. Uncertainty about responsibility of different Hqs/staff echelons.	a. Lack of clear division of labor between Div. & Bde. echelons.	a. Possible over-centralization at Div. level.	a. Lack of clear division of labor between Div. & Bde. echelons.	a. Bde. performed Div-wide planning functions.
I.B. Evaluation as ARTEP Tool	1. Doubts about frequency & timing of external evaluation.	a. Regular annual spacing of external evaluation exercises. "Everyone does it once a year."	a. Same.	a. Same	a. Unascertained.
		b. Evaluation date not coordinated w/ progress of specific Bn. training program	b. Same	b. Same	b. Unascertained.

* Note: The items reported in this table deal exclusively with observed *deficiencies* in the evaluation procedures of units selected for field study. In every case, such deficiencies should be interpreted against a predominantly positive background of effective procedures, eagerness to profit by experience, and enthusiasm for ARTEP as an innovative training/evaluation system.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
I.B. Evaluation as ARTEP Tool (Cont'd	2. Uncertainty about separating programmatically external from internal evaluations.	a. No reported use of Bn internal evaluations.	a. Same	a. Same	a. Evals. generally informal, blend of internal and external features. ("Just another training day – no distinction between the two.")
		b. Questions about distinction/difference between formal/external & informal/internal evaluations.	b. Same	b. Same	b. Not applicable. (Conscious and successful blend of internal and external aspects).
	3. Need for balance in emphasis & resources between evaluation & training activities.	a. Overall marginal support for evaluation process	a. Possible over-emphasis & over-spending on "formal" evaluations.	a. Some over-spending (e.g., POL)	a. Not applicable. (Need for balance well appreciated).
		b. Need more guidance on appropriate mix of evaluation & training activities.	b. Same	b. Same (Bde Commander stressed "big T, little E").	b. Not applicable.
	4. Test/report card tendencies. (Problems in maintaining the diagnostic perception/definition of the evaluation.)	a. "Report card" perception by Bn personnel, stemming from S/U ratings, competitive atmosphere. (Freq. refs. to "taking our ARTEP, rumors of relieved Bn C.O.s, etc.).	a. Same (High visibility & frequent interventions of Div. Commander encouraged pressure/test perception, competitive atmosphere).	a. Perceptions varied. Competitive atmosphere.	a. Unascertained.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
I.B. Evaluation as ARTEP Tool (Cont'd.)	4. (Cont'd.)	b. Lack of provision for detailed feedback. Apparent emphasis on pass/fail aspect.	b. Same	b. Same	b. Not applicable. (Very detailed feedback provided, followed by closely monitored remedial training).
		c. Occasional negotiation of ratings & arbitration during exercise (between evaluators & subject unit).	c. Same	c. Not applicable.	c. Unascertained.
		d. Apparent reluctance of evaluators to make neg. judgments.	d. Same	d. Possible over-criticality by outside CGSC evaluators.	d. Unascertained.
	5. Need for guidelines for blending evaluation & training objectives *during* the evaluation exercise (e.g., how much on-line feedback).	a. Lack systematic provision for on-line critique/instruction.	a. Same	a. Same	a. Unascertained.
		b. Interruption of scenario to accommodate performance repetitions.	b. Sr. Cmdr. terminating missions to inject critique.	b. At Co. team level, sporadic but intensive critiques *during* tactical phases.	b. Unascertained.
I.C. Unit Contextual Issues	1. Lack of sufficient representation of specific unit missions in Evaluation Plan.	a. Scenarios didn't reflect unit contingency missions.	a. Same	a. Same	a. Not applicable.
	2. Routinized uses of terrain in evaluation exercise.	a. Tactical situations required uses of overly familiar terrain in overly familiar fashion.	a. Same	a. Same	a. Not applicable. (Imaginative employment of rotating, three-lane approach to training area use).

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed *through Interviews*) Bns. D_1, D_2, D_3
I.C. Unit Contextual Issues (Cont'd.)	3. Plan design too ambitious for unit proficiency levels & evaluation experience.	a. Crowded schedule of events (too many missions), etc.	a. Same	a. Same	a. Same
		b. Need for greater expertise in control & tactical simulation of large units in complex missions.	b. Same (Frequent breakdowns in exercise control, etc.).	b. Same (Many unnecessary interventions by controllers).	b. Unobserved.
I.D. Task Assignment Issues	1. Centralization of evaluation preparation phase.	a. Not applicable.	a. One plan was used for all Div. Bns. Lack of consultations/ design refinements for Bns.	a. Not applicable.	a. Not applicable.
	2. Lack of inter-echelon coordination in preparatory phase.	a. Bde. Hq/staff provided with little division guidance.	a. Bde/Bn/Hq not involved in plan development.	a. Apparently little division guidance.	a. Little division guidance.
	3. Div. & Bde. roles not explicitly defined in execution phase.	a. Div. Comd./staff personnel played VIP observer roles, distracting Bde./Bn efforts.	a. Div. Cdr. exercised personal control w/o codified procedural guidelines.	a. Uncertainty about "play" of Bde. Hqrs.	a. Div. staff 0-6 served as nominal Sr. Evaluator. (Bde. Cdr. played this role, in fact.)
	4. Lack of fully spelled out responsibilities of Div./Bde/Bn in results utilization process.	a. Questions about timing and level of detail for feedback from Bde. to Div. and Bn.	a. No known Bde/Bn planning to use data to revise training programs.	a. Not applicable. Uncertainty about Div. Requirements for info. (& Div. future monitoring roles).	a. Unascertained.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
I.D. Task Assignment Issues (Cont'd.)	4. (Cont'd.)	b. Questions about Div. requirements for information.	b. Div. staff informal expressions of approval to Bn elements (unauthorize statements from Div- out of chain of command).	b. ADC chaired the critique w/ indeterminant but potentially significant (distorting) effects on the content.	b. Unascertained.
II.A. Exercise Format	1. Continuing use of 2 Bn (OPFOR) format.	a. Not applicable. (One Bn and OPFOR).	a. Division rqmt. based on belief in greater economy, realism, etc.	a. Bde requirement based on belief in greater economy, realism, etc.	Not applicable. Three unit round robin; elements played aggressor.
	2. Uncertainty about appropriate *composition* and exercise play of task force elements.	a. Poor simulation/non-representation of most support units.	a. Same	a. Same	a. Same
		b. Routine use of cross-attachment of Co. Team to opposite TF (Mech. Inf/Tank) or vice versa.	b. Same	b. Same	b. Not applicable. (Light infantry division).
	3. Lack of balance with respect to evaluation of separate Bn elements.	a. Overcommitment of evaluation personnel to Bn level.	a. Same	a. Not applicable. Balanced assignment of Evaluators, incl. 4 to each Company).	a. Unascertained.
	4. Sub-unit evaluations not coordinated with Bn. evaluation process.	a. Some sub-unit evals. conducted concurrent w/ Bn level evaluation but off-line, leaving holes in Bn & evaluation/control group.	a. Not applicable. (Sub-unit evaluations held before & after Bn eval.).	a. Not applicable; (sub-unit evaluations held before & after Bn eval.).	a. Many sub-unit evals. integrated into Bn scenario.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B. Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
II.A. Exercise Format (Cont'd.)	4. (Cont'd.)	b. Post-exercise subunit evaluations, (especially live fire), required performance by fatigued personnel.	b. Pre-exercise subunit evaluations, (conducted immediately prior to Bn. eval.), again tended to fatigue troops for Bn eval.	b. Same as Div B.	b. Not applicable.
II.B. Formulation of Evaluation Plan: Scenario Construction	1. Scenario over-structured due to complex evaluation requirements.	a. Frequent interventions in tactical situations to steer unit according to predetermined mission sequence.	a. Same	a. Same	a. Unascertained.
		b. High visibility and obtrusiveness of Evaluator/Controller Group personnel.	b. Same.	b. Same. Extremely.	b. Unascertained.
	2. Too many missions attempted in 3½ day exercise.	a. Time compression of individual missions unrealistic.	a. Same	a. Same	a. Unobserved, but probably same.
		b. Time compression of individual missions left little time for lower level troop leading procedures.	b. Same	b. Same	b. Unobserved, but probably same. (Bde Cdr suggested a 10 day exercise would be "ideal").
	3. Scenario predisclosed and routine in nature. (Lack of spontaneity and surprises.)	a. Unit was able to rehearse known scenario.	a. Same	a. Same	a. Same
		b. Absence of significant surprises or unforeseeable tactical development (e.g. OPFOR closely followed predetermined scenario).	b. Some surprises present due to 2 bn format and degree of "free play."	b. Same as Div. B. (Nuke and heli-borne raid played without opponent foreknowledge.)	b. Unascertained.

Table 3-1

(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
II.B. Formulation of Evaluation Plan: Scenario Construction (Cont'd.)	4. Intensity of activity uneven below Co. officer level.	a. Troops not consistently active (seldom dismounted, etc.)	a. Same	a. Same	a. Unascertained.
		b. Troop perceptions of exercise as "just another terrain ride"	b. Same	b. Same	b. Unascertained.
	5. Length of exercise encouraged unrealistic endurance requirements of leaders for training purposes.	a. Leaders went sleepless; no use of back-up leadership capabilities.	a. Same	a. Same	a. Unascertained, but presumably same.
II.C. Tactical Problem Control	1. Lack of support & planning of tactical simulation efforts.	a. Lack of realistic OPFOR force ratios.	a. Same (U.S. Bn *vs* U.S. Bn).	a. Same (U.S. Bn *vs* U.S. Bn).	a. Same (Reinforced Co as OPFOR).
		b. Degree of tactical *freeplay not fully* defined.	b. Same (Controller interventions not preplanned or guided).	b. Same	b. Unascertained.
		c. Underrepresentation or emphasis of unconventional warfare aspects. *(NBC, EW)*	c. Same	c. Same	c. Unascertained.
	2. Lack of clear distinction between (and procedures for) evaluation control requirements and simulation control requirements.	a. Too much "double-hatting" and overloading of E/C Group personnel.	a. Same	a. Same	a. Unascertained.
		b. Absence of detailed control plan and guides for simulation.	b. Same	b. Same	b. Not applicable. (Well developed control plan and simulation guides.)

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews Bns. D_1, D_2, D_3
II.C. Tactical Problem Control (Cont'd.)	3. OPFOR play unrealistic in both two-sided (due to nature of OPFORs) and one-sided (due to poor planning).	a. Need for better communication and control of OPFOR.	a. Same (U.S. tactics vs U.S. tactics, etc.)	a. Same	a. Same (as reported).
		b. Lack of Warsaw Pact type tactical doctrine, organization, equipment.	b. Same	b. Same	b. Same (as reported).
	4. Need for more adequate & timely play of weapon effects & signatures, casualties & equipment losses.	a. Lack of appropriate types, quantities of pyrotechnics.	a. Same	a. Same	a. Same
		b. Need for better communication/coordination to declare weapon effects/casualties.	b. Same	b. Same (although vehs. given REALTRAIN numbers to assist in control).	b. Not applicable. Well codified procedures using probability tables.
		c. Inconsistent declaration of casualties & duration of casualties' status.	c. Same (Roles for casualty assessment not explicit).	c. Same (Roles for casualty assessments not explicit).	c. See above.
		d. Inconsistent declaration of equipment losses.	d. Same (Evaluator decision with little known guidance).	d. Same (Evaluator decision with little known guidance).	d. See above.
	5. Need for better representation & utilization of staff & support elements.	a. S1/S4s roles underplayed.	a. Same	a. Same	a. Unascertained.
		b. Need more realistic processing of casualties, equipment losses, prisoners of war & intelligence.	b. Same	b. Same	b. Unascertained.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
III.A. E/C Group Attributes	1. Questions about organization of the E/C group.	a. Few evaluators at Co./Plt. levels.	a. Same	a. Four evaluators per company.	a. Same as Division C.
		b. Only one E/C group radio net.	b. Same	b. Not applicable, had two overlapping nets.	b. Same as Division C.
		c. Questions about chain of command in E/C group.	c. Same (role of Div Cdr not codified).	c. C&GSC students used with little orientation time.	c. Unascertained.
		d. No reserve evaluator/control capability.	d. Same	d. Same	d. Unascertained.
	2. Need better selection and training procedures of E/C group personnel. (See III.B., III.C., and III.D. below.)	a. Many evaluators from same parent unit (bde.), not recommended ranks.	a. Not applicable.	a. Not applicable. (C&GSC and other Bde provided E/Cs).	a. Same as Division A.
		b. Less than a single day for E/C training.	b. Same	b. Same	b. Unascertained.
	3. Materiel and support problems.	a. Evaluator vehicles had limited cross country mobility.	a. Same	a. Same	a. Not applicable.
		b. Insufficient radio operators, poorly maintained radios.	b. Same	b. Same	b. Unascertained.
		c. Inadequate pyrotechnics for safety & control purposes.	c. Same	c. Same	c. Unascertained.
		d. Dependency on Bde. for mess facilities, POL, etc.	d. Same (except Div. evaluation control center supplied support).	d. Same	d. Unascertained.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Areas of Apparent Deficiency
		(Applies to all cases)
III.B. Evaluation Criteria	1. Technical problems in mission T&E outlines (These are not related to performance of specific E/C Groups).	a. Lack of full correspondence between Task/Conditions/Standard elements.
		b. *Level of detail of coverage* of behaviors varies between missions.
		c. No space allowed for recording of many critical unit behaviors not anticipated in T&EO.
		d. High degree of dependence on "professional/subjective judgment" in interpretation of elements.
		e. Questionable correspondence between doctrinal literature and T&EOs.
		f. *Lack of adequate guidance* in coordination and integration of observations and ratings.

Table 3-1

(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
III.C. Evaluator Field Observation & Rating Procedures	1. Need for additional training in field use of T&EOs.	a. Less than half day's instruction on how to use T&EOs in the field.	a. Same (Evaluator school conducted in one day).	a. Same as Div. B.	a. Not ascertained.
		b. Need for instruction on use of "subjective" judgment elements.	b. Same	b. Same	b. Not ascertained.
		c. No joint terrain inspection.	c. Same	c. Same	c. Not applicable.
		d. No instruction regarding dual evaluator/controller functions.	d. Limited instruction	d. Limited instruction	d. Not applicable.
	2. Lack of preparedness to *anticipate* tactical developments as required to observe.	a. Lack of detailed gaming beforehand.	a. Same	a. Same	a. Not applicable.
		b. Few updating instructions to evaluators regarding tactical situation & desirable positioning.	b. Only limited updating.	b. Limited updating by a few Sr. Evaluators.	b. Unascertained.
		c. Evaluators reacting to rather than anticipating the action.	c. Same	c. Same	c. Unascertained.
	3. Need for more standardized format & procedural guidelines for note-taking & supplementary observations.	a. Some evaluators made few supplementary notes, some made extensive notes. Few E/Cs took T&EOs to field.	a. Same	a. Same	a. Unascertained.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
III.C. Evaluator Field Observation & Rating Procedures (Cont'd.)	3. (Cont'd.)	b. Many evaluators attempted to rely only on memory for later critiques & ratings.	b. Same	b. Same	b. Unascertained.
		c. Many evaluators expressed desire for much more detailed "checklists."	c. Same	c. Same	c. Same.
	4. Uncertainty about when to make ratings and when to communicate them.	a. Some evaluators made S/U ratings immediately; others waited until end of exercise.	a. Same	a. Same	a. Unascertained.
		b. Some evaluators conducted *outgoing* critiques; others supplied only completed T&EOs.	b. Same	b. Same	b. Unascertained.
	5. Limited coordination for optimal positioning to follow the flow of exercise.	a. Only a few senior evaluators cued their subordinates.	a. Same	a. Same	a. Unascertained.
		b. Most evaluators did not have sustained communication with OPFOR or detailed & timely sit. reps.	b. Same	b. Same	b. Unascertained.
	6. Uncertainty about how & when to integrate field observations & ratings.	a. Many evaluators rated behaviors w/o making necessary contact w/ other evaluators to integrate data.	a. Same	a. Same	a. Unascertained.

Table 3-1

(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
III.C. Evaluator Field Observation & Rating Procedures (Cont'd.)	6. (Cont'd.)	b. Lack of clear guidelines about integration across items/unit elements.	b. Same	b. Same	b. Same
IV.A. & B. Results Formulation and Communication	1. Need for more effective provision for comprehensive post-exercise verbal critique.	a. Lack of complete attendance by E/C group & evaluated unit leaders.	a. Not applicable.	a. Same as Div. A.	a. Unascertained.
		b. Lack of detailed agenda.	b. Not applicable.	b. Same as Div. A; session chaired by ADC.	b. Unascertained
	2. Lack of adequate provision for post-exercise E/C group joint discussion to integrate observations & ratings.	a. Group meeting focused on providing "impressions" to Sr. Evaluator.	a. Unascertained.	a. No coordinating meeting held.	a. Unascertained
		b. Pressure on evaluators to adjust ratings to bring into line with Sr. Evaluators ratings.	b. Unascertained	b. No coordinating meeting held.	b. Unascertained.
	3. Need to provide timely & valid feedback at troop level.	a. Lack of detailed post-exercise critiques for troops.	a. Same	a. Not observed.	a. Same
		b. Feedback was delivered in widely varying forms with widely varying effectiveness.	b. Same	b. Not observed.	b. Unascertained.

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
IV.A. & B. Results Formulation & Communication (Cont'd.)	4. Written critique lacked detail, did not pinpoint weaknesses.	a. Content focused on overall judgments supplemented by only a few "strengths and weaknesses"	a. Same	a. Unascertained.	a. Unascertained.
	5. Uncertainty regarding form, content, audiences and timing of formal written evaluation report.	a. Bde. Cdr. as St. Evaluator regarded himself as the chief audience.	a. Not applicable. (Clear reporting guidelines provided).	a. Bde Cdr, as Sponsor, defined reporting requirements to meet own needs.	a. Not applicable. (Clear reporting guidelines provided).
		b. Written report to Div. was highly summary in content.	b. Div., as sponsor, required relatively detailed reporting.	b. Same as Div. A.	b. Same as Div. A.
	6. No experimentation w/alternative (e.g., graphic/electronic) media to enhance feedback processes.	a. Only verbal/written media used.	a. Same	a. Same	a. No alternative media reported.
IV.C. Results Utilization	1. Lack of formal specification of responsibilities of Command/Staff echelons in remedial training process.	a. No definite assignment of specific roles to Bn/Bde/Div Cdrs.	a. Same	a. Same	a. Not applicable. (Roles of Bde. & Bn. Cdrs. clearly defined).
		b. No detailed development of types & sequences of action with respect to results.	b. Same	b. Same	b. Not applicable.
	2. Lack of adequate provision for necessary blocks of remedial training time for full Bn.	a. Time constraints prohibited retraining of Bns as integral units.	a. Same	a. Same	a. Same

Table 3-1
(Continued)

Issue Area	Critique: General Comments	Critique: Specific Observations for Individual Evaluated Units			
		Div. A Bns. A_1, A_2	Div. B Bns. B_1, B_2	Div. C Bns. C_1, C_2	Div. D (Reconstructed through Interviews) Bns. D_1, D_2, D_3
IV.C. Results Utilization (Cont'd.)	2. (Cont'd.)	b. Apparent lack of emphasis for above requirement.	b. Same	b. Same	b. Unascertained.
	3. Tendency to ignore results implications if unit "passed."	a. Apparent letdown tendencies present in evaluated units.	a. Same (Apparent overemphasis of S/U ratings).	a. Unascertained.	a. Not applicable. (Cdrs. reportedly stressed remedial training requirements).
		b. Apparent lack of adequate command emphasis.	b. Same	b. Unascertained.	b. Not applicable.
	4. Need for systematic monitoring of remedial training by Bde.	a. Broad guidance primarily embodied in the formal report, apparently not followed by detailed inspection/observation.	a. Insufficient info. G-3 reported scheduling remedial training. Monitoring was not observed.	a. Not applicable.	a. Not applicable.
	5. Need for effort to coordinate training at differing unit levels to optimize effectiveness of evaluation data for entire Bn.	a. Tendency to be distracted by "other" missions & then to resume "cycled" training at lowest level.	a. Same (as reported in follow-up contacts).	a. Same (as reported in follow-up contacts).	a. Not applicable.
	6. Need for systematic follow-up effort to review evaluation experience of E/Cs for future refinement of evaluation procedures.	a. Major overhauls of evaluation procedure reported, but how influenced by this experience not known.	a. Limited use reported.	a. Unascertained.	a. Not applicable.

under I.A.2.a in Table 3-1. It did not apply in the case of Division D where elaborately codified definitions of specific Standard items were the rule. An alternative solution, which does not assume we can greatly reduce the subjectivity in judgments, involves training evaluators to exercise wise subjective judgments.

At the pay-off end of the evaluation process, it appeared that division training schedules were too fixed and inflexible to make more than nominal use of results of ARTEP evaluations. Typically the master schedules are prepared a year in advance, with "prime time" training periods scheduled for each battalion only once in that interval. This perpetuates the peaking phenomenon, and makes full battalion remedial exercises exceedingly hard to squeeze in. In all cases observed and shown under I.A.3.a in Table 3-1, there was little evidence of sustained battalion level post-evaluation training. The peaking tendencies in the overall T&E program may well be the most essential factor under this heading because it involves the integration of battalion field exercises with all other forms of training.

As a last major problem, there appeared to be some confusion as to procedures for establishing effective coordination and a clear division of labor between division, brigade and battalion echelons during various phases of the evaluation effort. Our research team did not actually observe the formation of an "ARTEP Committee" nor the planning phase of an external evaluation exercise. Rather our observations are based upon interviews conducted the day before the inception of an exercise with staff personnel involved in the planning stage. Results of these interviews are presented under I.A.4.a in Table 3-1. As extreme examples, individual brigades either performed division-wide planning functions, or operated with virtually no help from division in any phase.

From these observations, it is apparent that some of the most central philosphical principles of the ARTEP system are not effectively grasped in field units. On a somewhat less abstract level, similar types of issues are detailed in the next section.

(2) Using Evaluation as an ARTEP Tool. Five major problem areas were identified with respect to the use of evaluations as tools within the context of the complete ARTEP system.

(a) Uncertainty about when to conduct the various types of evaluations.

(b) Problems experienced in distinguishing between the various types of evaluations (e.g. internal/external, formal/informal).

(c) Problems in attempting to establish the correct mix of alternative training and evaluation options.

(d) Problems in using the external evaluation as a training and diagnostic tool rather than as a report card.

(e) Uncertainty about how to balance evaluation and training activities and objectives "on-line," i.e. during the actual battalion evaluation exercises.

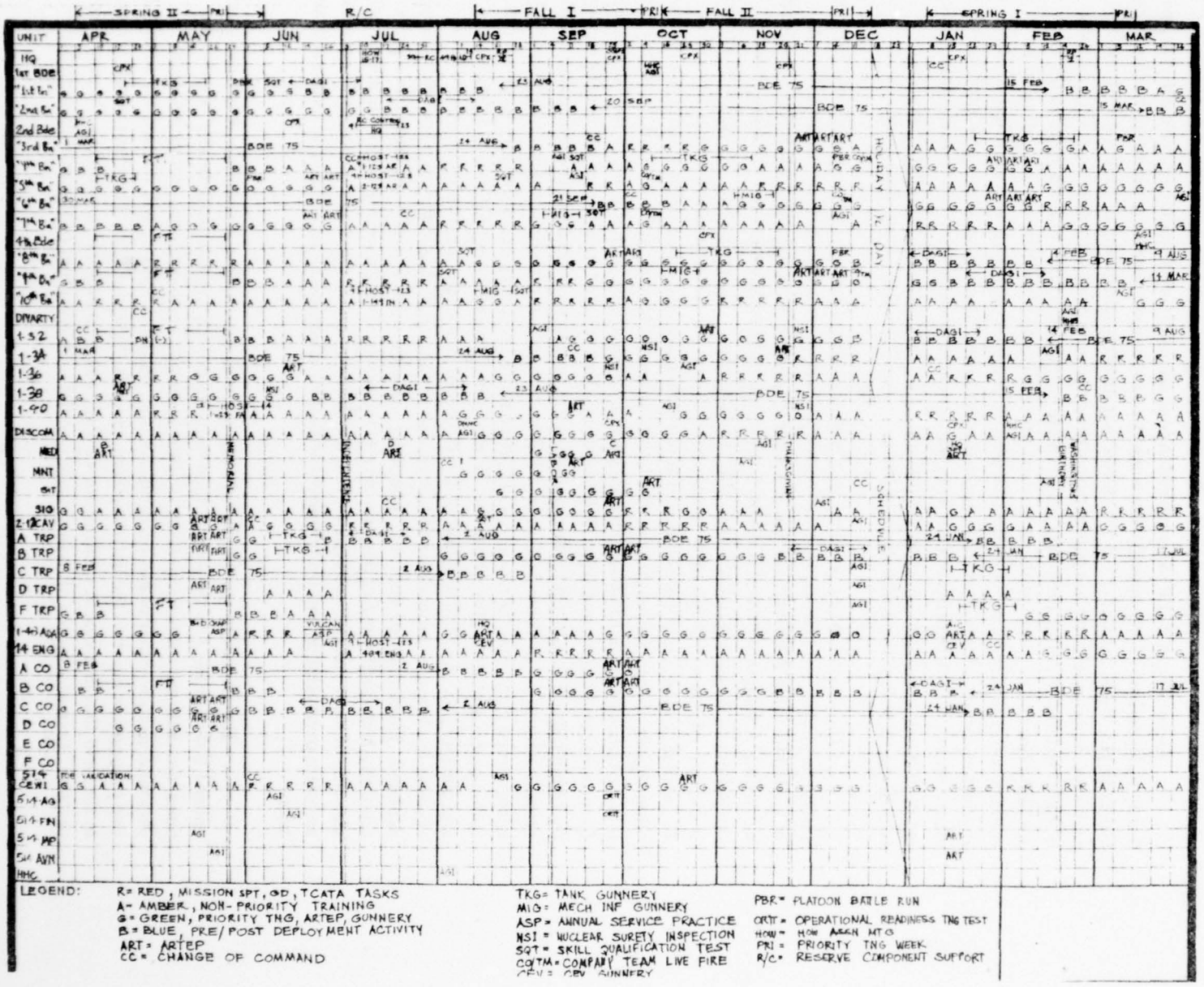
Generally, ARTEP external evaluations are routinely scheduled once each year, evidently independent of any specific assessments of unit needs, level of evaluation readiness, etc. For example, Table 3-1 under I.B.1.b shows that division scheduling of external evaluations for each specific brigade/battalion was apparently not coordinated with the progress of specific brigade/battalion training programs within these annual intervals. In both aspects, the core ARTEP principle of tailoring the evaluation process to meet the particular needs of units collides with the very heavy coordinating and scheduling constraints that confront divisions and higher command levels, as they juggle access to training areas, multiple mission commitments, and so on. A Division Master Training Schedule documenting the typical annual spacing of external evaluations is presented in Table 3-2. Inspection of this chart reveals only two elements within the entire division performed an "ARTEP" more than once a year.

The second major problem area is closely linked with the first. Many field personnel interviewed were confused as to how the ARTEP approach really differed from ATTs. As indicated in Table 3-1 under I.B.2.a, few sampled units conducted any total internal evaluation activities. A representative of TRADOC suggested internal battalion field exercises be conducted quarterly for several reasons. One, to maintain proficiency; two, because ammunition is distributed quarterly; and three, personnel instability dictates regular internal evaluations. In parallel fashion, units were confused about how much and when to do which type of evaluation. Partly due to this confusion, field units appeared to have little basis for deciding how much emphasis to give to battalion-sized external evaluations, and how much to give (spend) for internal evaluations and other T/E activities. This resulted in apparent over-spending and over-emphasis on formal (external) evaluations in several units, and under-emphasis in several others as indicated in Table 3-1, under I.B.3.a. One sergeant major summarized a frequently encountered view, that battalion ARTEP evaluations are a very expensive way to check out commanders and staffs, for whom the use of simulation games and CPXs could do a better and cheaper job. As an illustration of the effects of over-spending, following one ARTEP exercise the brigade commander conceded that his annual POL allowances were completely depleted. Division D, by contrast, reportedly made do with relatively few assets to achieve relatively high quality evaluations (largely as the result of very thorough planning and understanding of basic ARTEP concepts). Taken together, the limited data now available tend to associate overspending and overemphasis with division-centered approaches to external evaluation.

The evidenced confusion between internal and external evaluations bears on issues under I.B.4 in Table 3-1. So long as field units do not distinguish procedurally between internal and external evaluations there is always going to be some doubt among battalion personnel as to whether or not performance on a battalion ARTEP will be used for administrative purposes as input to the battalion commander's "report card." There was ample evidence in the field that many units continue to view the external evaluation as more a "report card" or "test" than as a source of usable diagnostic data. A symptom of this mistaken perceptio

TABLE 3-2

REPRESENTATIVE DIVISION MASTER TRAINING SCHEDULE



was a lack of provision in the evaluation plan for detailed feedback procedures. (As indicated in Table 3-1, under I.B.4.b, Division D was again an exception; its plans were well developed.) In general, there was a tendency to focus on the T&E sat/unsat ratings, rather than on detailed critical notations. In several divisions, the atmosphere at brigade and battalion levels was characterized by intense competitive pressures, rumors of poorly performing battalion commanders being relieved, and so on. Commanders and troops still talked about "taking ARTEPs" as one takes a test. In this sense, they may be said to have been more interested in achieving "S's" and avoiding "U's" than receiving frank, detailed feedback about unit strengths and weaknesses. This defensive attitude sometimes led to efforts at negotiation between commanders and evaluators over rulings and ratings; and to some reluctance by evaluators to make negative judgments. Members of our research team discussed this problem of dual usage with a senior officer at TRADOC. He suggested that the two functions, i.e. training and accountability reckoning, be maintained but separated procedurally by conducting two battalion ARTEPs annually. One field exercise would be primarily for training diagnosis which presumably would preceed the other for testing. In this way the Army might increase the credibility of the use of field exercises for "sanction free" training diagnosis. However, feasibility questions remain due to serious constraints on resources.

Finally, field observers repeatedly recorded doubts over how to deal with training and feedback activities while the battalion evaluation was in progress. A primary source of this confusion was the lack of definitive policy and procedural guidelines in the evaluation plan for the administration of on-line critiques and/or instruction. As a consequence, several types of uncoordinated training activities were observed which tended to undermine the realism and the logical flow of the exercises. For example, issue area I.B.5.b in Table 3-1 shows that in a number of cases, missions were interrupted "prematurely" or arbitarily for immediate performance critiques, instruction, or task/mission repetitions.

(3) Unit Contextual Issues. One major facet of ARTEP involves tailoring T/E efforts to match particular unit's needs and capabilities. Several types of problems were identified that relate to this process. Among the most important were:

(a) Development of the Evaluation Plan, particularly in the selection and tactical definition of missions, which did not fully reflect the distinctive configuration of TO&E and contingency missions for the units being evaluated.

(b) Less than full and imaginative exploitation of available terrain for conduct of the evaluation problem, chiefly by recourse to employment of overly familiar terrain in routinized fashion.

(c) Development and attempted use of overly ambitious Evaluation Plans requiring unrealistic levels of sophistication/proficiency for both evaluator groups and evaluated units to execute effectively.

By and large scenarios did not reflect the particular mix of TO&E missions, operational contingency missions, "ALO" numbers, and so on, which define the probabilities for actual combat situations that each individual unit faces. The training doctrinal literature (e.g. TC 21-5-7) mandates this as the starting point for design of a unit T/E program. Instead, Evaluation Plans and scenarios were virtually interchangeable across our entire sample of Tank/Mechanized Infantry task forces.

Terrain limitations are often such that with firing ranges in use and other training activities, it is difficult to find sufficient room for tank/mechanized battalions to maneuver.

Fresh thought as to how to get maximum evaluation exercise benefits from imaginative use of familiar terrain is warranted. Particularly at Division B there was a pronounced pattern of adopting stale terrain situations and usages for employment in the evaluation exercises. Although obviously there are real limits to the flexible use of training areas, repetition of the same missions over the same ground does little to inject surprise, fresh tactical thinking, or troop enthusiasm into the ARTEP exercise.

Finally—and this was universal—Evaluation Plans called for much higher levels of evaluator/controller expertise and proficiency by the unit evaluated than were present. The third issue area under I.C. in Table 3-1 interacts with the previous issue in so far as evaluation plans contained too many missions which were unrelated to operational, real world contingency missions. We recommend simpler, less ambitious scenarios and formatting. This would help eliminate frequent lapses in exercise control and realism, and situations where units were called upon to execute complex tactical exercises when the inability to execute fundamentals is questionable. In such cases, scarce resources were being misallocated due to readily preventable causes.

(4) Evaluation Task Assignment Issues. ARTEP and associated training documents call for high levels of decentralization in all aspects of unit training and evaluation procedure. However, uncertainty concerning how this should be interpreted in the assignment of evaluation roles to participating division/brigade/battalion command and staff elements was inferred by HSR researchers from conversations with the personnel involved in the planning stage.

In one case, shown in Table 3-1 under I.D.1.a, the effort was completely centralized at the division level. A single Evaluation Plan was created and used for all battalions in that division. Obviously, efforts to tailor the process to specific battalion circumstances were minimal in this case. The same may be said for the level of participation and planning exerience gained by brigade/battalion personnel. Division C was a clear exception to the division centralized task organization of the ARTEP Planning Committee. Pre-exercise interviews with the G-3 revealed that division attempted to increase the participation of brigade personnel in the preparation phase by tasking the Assistant Brigade S-3 with designating the Evaluation Plan. The plan was then distributed to selected staff and commanders for inputs to refine the design. The plan and solicited inputs were then sent to the G-3 for final approval.

In some instances, virtually no direction from the division was forthcoming in the preparatory, execution or feedback phases of the complete evaluation process. In all cases observed and reported under I.D.3.a in Table 3-1, at least some of the leadership and staff talent at the division/brigade echelons was not provided a meaningful role in the field. This resulted in brigade commanders drifting in and out of the scene, looking over their battalion commander's shoulder, and distracting the efforts of others because commanders were "being simulated" by somebody else in the exercise plan. We feel that frequently senior commanders gave too much attention to the battalion being evaluated; too little to the effective operation of the E/C Group.

Similarly, there was substantial ambiguity as to who was responsible for what in feeding back evaluation data to remedial training practices. This problem was aggravated by the fact that Evaluation Plans put out by the sponsoring divisions/brigades did not specify the roles or tasks of various division/brigade personnel in utilizing the evaluation results. Again to cite extreme cases, brigade commanders who had to mount the entire effort virtually alone were reluctant to let division people dabble with how the results were to be utilized. They preferred instead to supply division with minimal reports and information.

b. Evaluation Exercise Preparation and Conduct Issues

(1) Evaluation Exercise Format. The major problem areas identified with respect to basic Evaluation Plan design decisions were:

(a) Use of two-battalion ("OPFOR") exercise formats *without full awareness of the deficiencies and difficulties of this approach.*

(b) Confusion about the appropriate logic and procedures for determining the composition of the performing task force and supporting units.

(c) Over-concentration of evaluation assets at the battalion level, with consequent losses of key data at the lower unit levels.

(d) Uncertainty regarding how to coordinate battalion and sub-unit evaluation procedures.

Despite FORSCOM and TRADOC guidance, several divisions continued to employ the two-battalion ("OPFOR") format in evaluation exercises as shown in Table 3-1 under II.A.1.a. They remain outspoken in defending this approach, based on the highly questionable view that it is a more practical, cost-effective, and realistic format. Having given considerable attention to this issue, we believe that the single battalion format is more easily managed, permits more authentic representation of Warsaw Pact-type OPFOR elements, allows for a greater degree of exercise control, and is subject to less costly and more rigorous evaluation. Therefore, we believe it is to be preferred.

There was a great deal of uncertainty regarding the number and types of supporting elements to put in the field, or simulate. In many instances, critical combat and service support elements were either poorly simulated or omitted from the exercise altogether, thus detracting from the realism of the exercise. On what is perhaps a picayune further point, we were inclined to question the unreflecting and mechanical way in which units created Tank/Mechanized Infantry Task Forces by *always* simply swapping company-sized elements. The "How to Fight" Manuals emphasize flexibility in tailoring Task Forces to specific tactical settings.

Another common problem was questionable balance in the distribution of evaluation assets between battalion elements. There was a tendency to assign a disproportionate number of evaluation personnel to observation of various aspects of battalion-level activities. As a result, mission-critical performance dimensions observable only at the lower unit levels were inadequately covered.

Finally, there was a wide divergence in approaches to the conduct of "sub-unit" evaluations, with some less successful than others. In several cases listed in Table 3-1 under II.A.4.a, parent units attempted to conduct the bulk of these evaluations concurrent with the battalion evaluation. This stretched evaluator resources very thinly and left unrealistic gaps in battalion formations. One G-3 concluded that concurrent sub-unit evaluations detract from participants' perspective of the whole. In other cases, listed under II.A.4.b, "off-line" approaches were not scheduled so as to avoid troop fatigue, either during the battalion exercise or the sub-unit evaluations.

(2) Evaluation Plan/Scenario Construction. Five common weaknesses of the Evaluation Plans and associated scenarios were identified.

(a) Commitment to *overly complex evaluation requirements* resulted in unwieldy exercise scenarios. (In order to make these "work," units were often given advance copies of scenarios and mission schedules.)

(b) The *universal attempt to cram an excessive number of missions* into the 90 hour (+) exercises.

(c) Prior disclosure of the detailed scenario, and similar compromises of tactical uncertainty, etc.

(d) Plan/scenarios which kept leaders very busy but often failed to exercise troops adequately.

(e) Plans which permitted "endurance test" performance by first-line leaders without requiring recourse to back-up leadership capabilities.

In general, exercises contained too many complicated combinations of missions compressed into too short a period of time. As a result, performing units were forced to attempt to complete missions within an unrealistically short period of time. Not infrequently there was no time left for subordinate elements to practice troop leading procedures. For the same reason, it was frequently necessary for controllers to intervene and adjust tactical situations in relation to the fore-ordained schedule of events. Additional reasons for interventions were lack of planning by the E/C Group so as to *anticipate* battalion actions. This undoubtedly resulted from an insufficient amount of pre-exercise time spent training and assigning roles to evaluators and controllers. A further consequence often observed was the high visibility and obtrusiveness of E/C personnel in the field. But this varied.

In some cases evaluators did not stay with their units and could not be found. In others they stationed themselves (or rode around in jeeps) just in front of units being evaluated, thus helping to expose their positions to the opposing force. The compression of missions also tended to produce, by direct and indirect causes, overburdening of unit leaders from company up and under-utilization of troops and back-up leader capabilities. While troops typically were active only sporadically, (and frequently perceived the exercise as "just another terrain ride") company level and higher leaders often went sleepless for the duration of the exercise.

The Evaluation Plan/Scenarios also tended to take on "canned" or predetermined characteristics. Unit leaders were almost always aware of the sequence of missions and events in advance, and generally were able to rehearse before the evaluation was implemented. Reportedly in Division A, company level units did rehearse in anticipation of the known scenario. Nowhere did we find this practice discouraged. In some units, this serious compromise of surprise, realism and spontaneity was defended as a virtue. The logic was "the more the unit knows in advance, the better it can get ready." Of course, there were some surprises in the two-sided exercises. Too often E/C personnel were unprepared for what was going to happen next, thus being in the wake of events, they were unable to avert the muzzle-to-muzzle confrontations which occurred.

(3) Tactical Problem Control. The exercise of adequate control while achieving adequate degrees of tactical realism represented a critical deficiency, and one that is not readily remedied. Among the most serious shortcomings were:

(a) Both a lack of *adequate support* and a lack of *sufficient planning* for *tactical simulation.*

(b) *Failure to adequately distinguish and prescribe evaluation and tactical control simulation procedures.*

(c) In some instances, unrealistic employment of OPFOR elements.

(d) Inadequate and untimely play of weapons effects and signatures, as well as declaration of casualties and equipment losses.

Serious shortcomings were the rule with respect to materiel and personnel commitments in support of control and tactical simulation efforts. No units had available adequate pyrotechnics even to begin to approach "realism" in the simulation of weapons' firings and impact signatures. Perhaps more significantly, Evaluation/Control Plans typically failed to spell-out sufficiently well how weapons effects and force ratios were to be represented. The same was true of tactical "arbitration" and "umpiring" rulings in every exercise observed. A comparison of evaluation/control plans across units revealed no discernable distinction between one and two-battalion exercises. We believe the two formats require different control measures. As for personnel, the duties of the E/C personnel were not well defined, nor were there, to our knowledge, any rules or guides for shifting personnel to the scenes of hottest action. As a result, control often faltered, units went "off the board" or became entangled muzzle-to-muzzle. Much potentially useful training time was wasted in sorting people out and re-establishing the tactical flow.

In addition, we found universal problems with the size, composition, and tactical plan of opposing forces in the exercises observed. Such problems as realistic force ratios, simulating Warsaw Pact OPFOR tactics, etc., are unsolvable in the two-battalion format. The OPFOR, on the other hand, represents a separate and more flexible evaluation instrument. For example, Division A could have set up the OPFOR to play Soviet tactics. However, this area appears to receive little attention.

Finally, few units provided for realistic play of staff and support elements. A closely related deficiency was a lack of detailed follow-through on the processing of casualties, equipment losses, POWs and raw intelligence. Division C was a clear exception in that they attempted to integrate REALTRAIN identification techniques with their control measures. However, there was no net control station to monitor and record inflicted casualties as there is in REALTRAIN, hence the designation in Table 3-1, under II.C.4.b. The same rationale holds for II.C.4.c and d. Although REALTRAIN numbers were placed on tanks, use of those numbers in the declaration of casualties and equipment losses was inconsistent.

c. Evaluation Methodology

(1) Evaluator/Controller Group Attributes. Key deficiencies of E/C Group organization and training were:

(a) Problems of Group organization and procedure.

(b) Deficiencies in personnel selection and training procedures.

(c) Shortcomings in the nature and degree of support for the E/C Group.

There were serious organizational problems. In our view, personnel assignments were not clearly defined or coordinated. "Double-hatting" often forced E/C personnel to neglect critical duties (either in evaluation or simulation). We were sometimes confused (though Army

participants may not have been) as to command structure and operational control functions. Linkage with OPFOR elements was often weak and ill-defined. There was no confusion as to who was in command during Division C's exercise because the brigade commander was established as the Test Control Directorate. This Division also proved to be an exception at the other end of the E/C Group organization. In most cases, there were too few evaluators assigned to company and/or platoon units. Yet interviews with a G-3, who had participated in the planning stage of several battalion exercises, revealed a strong preference for decreasing or eliminating the number of evaluators assigned to performing platoons.

Other problems were a general lack of provision for an E/C Group reserve capability for evaluating attached units. (This problem is quite distinct from that of coordinating on-line assignments of evaluators amongst maneuver elements.) There were repeated communications breakdowns due to the absence of an adequate number of operable radios and group nets. Too frequently, the radios provided turned out to be inoperable in the field. As indicated in Table 3-1 under III.A.1.b, most units provided only one radio net/frequency to the entire E/C Group. An exception was Division C's radio net system. Two overlapping nets were used, one at battalion which could be heard by all E/C personnel and one at company over which transmissions could not be heard at battalion.

Training procedures for the E/C Group were uniformly found to be inadequate. Personnel were very frequently detailed from neighboring battalions, and were not of the prescribed rank as indicated in Table 3-1 under III.A.2.a. Lack of adherence to guidance provided in 71-2 regarding the appropriate rank of evaluators may have decreased credibility of feedback. However, we suspect that the problem of relative ranks between evaluator and evaluatee can be largely reduced by evaluator training. Much more harmful to the overall evaluation effort was neglect of the minimum requirements for training personnel for their evaluation and control duties. Training observed fell far short of that recommended in 71-2.

Technical support of the E/C team was often deficient, primarily in the areas of transportation and communication as shown in Table 3-1 under III.A.3.

(2) Evaluation Criteria. A considerable number of "technical" problems were identified which are inherent in the current 71-2 T&E outlines. These did not reflect shortcomings on the part of specific units. These may be roughly classified as:

(a) Apparent lack of correspondence between Task, Conditions and Standards components.

(b) Questionable linkage between doctrinal literature and the T&EO mission contents.

(c) Apparent lack of provision for critical events and behaviors in many mission outlines.

(d) Variations in the length and intensity of treatment of different missions.

(e) Lack of procedural guidelines for integrating observations and ratings across standards items, missions and unit echelons.

One of the major problems arising in the T/E outlines concerns forking/branching in the tasks, standards, and conditions. We found no evidence that the potential forks within T/E mission outlines, such as appears in Chapter 2, Part Two of the second volume of this report, had been analysed during the planning stage of exercises. Another anomaly became apparent when T/E items were compared between echelons for three major missions. Detailed results and implications of the comparative analysis are presented in Chapter 2 of the second volume. Briefly, there was a high degree of variance in the intensity with which behaviors were covered across missions. Observation III.B.1.c. in Table 3-1 refers to critical behaviors not included in the T/E outlines which require anticipation on the part of E/C personnel.

Another pressing issue is the lack of precision in the specification of Standards. This problem—as applied to ground warfare—is most complex. See discussion in Volume 2. At the present time, precise and valid objective numerical standards cannot be practically applied by operating units to *cover the bulk of mission duties* required in tactical field exercises. Subjective, professional judgments will continue to play a critical role until the state-of-the-art for unit standards is significantly advanced. Engagement simulation can help in this advance, but in and of itself will not solve the standards problem. New approaches such as gaming and Delphi methods should be explored as should the training of evaluators. Follow-on work must, therefore, provide further guidance bearing on the interpretation of T&EO Standards.

(3) & (4) Field Observation and Rating Procedures. Inadequacies in evaluator training have been mentioned repeatedly above. Specific manifestations included:

(a) Failure to delineate *all* evaluator functions and train in these.

(b) Little instruction as to how to use T&E outlines in the field, and exercise subjective judgments. As a result, field use of T&E formats varied widely.

(c) Lack of instructions as to how to evaluate performance requiring coordination between tactical elements not in visual contact.

(d) Lack of guidance for integration of observations and ratings made by different evaluators.

(e) Lack of guidance for handling field critiques.

(f) Finally, a lack of appreciation of the complexity and difficulty of evaluator tasks as individuals, and as a team.

In practice, E/C personnel have about five or six duties, of which evaluation and control are the most prominent. Others include providing feedback, insuring that safety rules are observed, and making records of performance. Typically, little instruction was provided on the use of T&E outlines in the field. As indicated in Table 3-1 under III.C.1.d., there was limited instruction on how E/C Group members should combine dual evaluator/controller functions and responsibilities. Nor did they receive or realize the importance of training in all essential evaluator/controller duties, as evidenced by responses to our questionnaire. Training did not normally include instruction in how to make subjective judgments and interpretations. Nor were evaluators trained in how to handle interdependent observations and ratings. We were unaware of any instructions provided evaluators as to who actually observes and rates what behaviors or events. This may appear simplistic at first glance, but when there are *identical* items across battalion, company and platoon mission outlines, this question arises: Is the rating for a battalion simply a summation of ratings made by evaluators with platoons, *or* are *independent* estimates made by evaluators at company and battalion levels? We were unable to find guidance on such questions in the instructions provided most E/C Groups.

Most units failed to conduct E/C Group terrain inspections. Thus, evaluator/controllers were not (well) oriented as to the likely flow of events and how it may bear on their own duties. As a result, most evaluators were put into the position of reacting to ongoing events rather than anticipating them. Had E/C Group training included a classroom gaming of the exercise as called for in Chapter V of ARTEP 71-2, problems might have been obviated. Once the problem was underway, only a few evaluators tried to overcome this handicap by cueing their subordinates for positioning and what to observe. As indicated in Table 3-1, under III.C.2.b., a few trained evaluators updated instructions to the evaluators assigned to them, and provided on-line critiques. The majority of E/C Group members did not possess continuous communications with the OPFOR. Clearly, procedural guidelines for adequate communications between the OPFOR and E/C Group are needed.

Our data also suggest that evaluators need a single standardized format and a set of procedural guidelines for note-taking and for recording supplementary observations. Recording behavior varied from extensive note-taking during the exercise to no record keeping at all. Many evaluators depended primarily on memory until breaks between missions, when they attempted to compile permanent data records. This method increases the likelihood of forgetting. If evaluators forget, they cannot perform their part in closing the learning loop. We believe that stress in evaluator school on observing/recording performance when it occurs, plus improved formatting of T&E outlines would result in more comprehensive and more valid descriptions of field performance. Thus, capabilities for training diagnosis would be much improved.

Finally, E/C personnel were uncertain about how and when to make actual rating entries in the T&EOs. They were also unclear about how and when to integrate field observations and ratings with other E/C Group members. Some evaluators formulated and recorded S/U ratings immediately, while others waited until the end of particular tasks or missions to do so. Many evaluators attempted to formulate ratings independently, even where coordination with other E/C was necessary. There was the same degree of variability in procedures

for communicating ratings and other data to the performing units. Few evaluators routinely conducted on-line critiques at the completion of each problem or mission. Others did little or nothing more than simply hand over copies of their completed T&EO forms to unit commanders. One evaluator went one step further than most of his counterparts during Division C's exercise. When his platoon was relatively inactive, he set up hypothetical situations for all to discuss. Exercises observed provided many opportunities for this type of "on-line" training.

d. Results Formulation, Communication and Use

(1) & (2) Results Formulation and Communication. There were deficiencies in formulating and communicating evaluation results. Among these were:

(a) Lack of realistic provision for post-exercise verbal critiques at all appropriate unit levels.

(b) In some cases, inadequate provision for bringing the entire E/C Group together after the exercise for the purpose of coordinating observations and integrating ratings.

(c) Lack of systematic provision in the LOI for timely and valid feedback at the troop level administered between missions *during* the exercise.

(d) Lack of sufficient detail and precision in the "formal"/written evaluation reports to pinpoint important performance deficiencies requiring remedial training.

(e) Uncertainty regarding the correct form, content, audiences and timing of the Formal Report(s).

(f) No attention to using alternative media to enhance the feedback process (e.g., graphics or electronic technologies).

As indicated previously under Issue Area a.(2)(e), a variety of feedback forms were observed in the field. Some of these entailed unplanned training activities such as task/mission repetitions. This issue area addresses deficiencies in three types of feedback and critique activities: on-line verbal critiques conducted during the exercise; post-exercise verbal critiques; and the written formal evaluation reports.

The Evaluation Plans which were distributed by sampled field units were closely inspected. They indicated a clear need for more guidance on procedures for administering on-line and post-exercise critiques. Such questions as whether, how and when to intervene and provide feedback during the conduct of a mission were not addressed in published evaluation plans provided us. Feedback, when provided, tended to be fragmentary. We observed no direct critiques conducted for unit enlisted personnel except in the case of Division C. Reportedly in Division C some evaluators attempted to informally critique their units following the field

exercise. Division C was the only sampled unit to provide written guidance with respect to who was to attend a post-exercise verbal critique. Here, better ways must be found to analyse errors objectively, while maintaining respect for commanders, and for the authority structure.

A coordination meeting for integration of observations and ratings was held by two divisions; we were unable to establish whether it was held by other divisions. There appeared to be a lack of uniformity across divisions as to who was to attend.

With the exception of Division D, the formal written evaluation reports were inadequately detailed for diagnostic purposes. Division D was the only unit which made a conscientious effort to pinpoint weaknesses. In a follow-up phone interview with the G-3 of one division, we were informed that the division had substituted for S/U ratings detailed notations of strengths and weaknesses. They had formerly used S/U ratings but believed this change will relax the apprehension observed among evaluated battalion commanders.

Guidance provided with respect to the form, content, audiences and timing of formal written evaluation reports varied between cases. One division appended a sample evaluation report format to their Evaluation Plan and emphasized the diagnostic utility of the results. One division provided detailed exemplars of battalion and company-written evaluation reports in their published Evaluation Plan.

From these observations it is apparent that field units need more systematic guidance on how and when to formulate and communicate feedback, and to whom.

(3) Use of Results. The research team had but very limited opportunity to observe use of evaluation results. Therefore, we attempted to determine by phone and correspondence whether and how results were being used for remedial training. Several types of major inconsistencies with ARTEP prescriptions were identified. These include:

(a) Lack of precision in the definition of planning, execution and monitoring functions for the multiple headquarter/staff echelons involved.

(b) A general inability to bring together enough resources to permit remedial training for the battalions as a whole.

(c) Limited efforts to compensate for the above by developing concurrent, multi-echelon training programs which are keyed to documented weaknesses.

(d) Need to review the evaluation process from the perspective of its future refinement.

The lack of formal specification of training managers' roles in the results use phase has already been touched upon under the first issue area a.(1)(d). We know of only one case in which the battalion commander was held responsible for incorporating his unit's evaluation results into the battalion Planning Calendar. Specifically, Division D's brigade commander required his battalion commanders to justify their revised plans in light of identified deficiencies.

The second major problem area involves the lack of flexibility to incorporate remedial training needs in master training schedules. This is an area requiring further study. We need to determine to what extent existing training schedules can be modified to use evaluation results effectively. As indicated in Table 3-1 under IV.C.2.A., most units were unable to schedule remedial field training for battalions due to time and training area constraints. We were informed that Division A has recently taken commendable steps to obviate this deficiency. They have revised their annual training schedule to enable battalions to return to the field within a month following an external evaluation. Alternative means of retraining were not fully explored by most other units when it was impossible to rerun a battalion as a whole (see IV.C.5. in Table 3-1). Division B was an exception to this general observation. We were informed that they scheduled two weeks of multi-echelon make-up training such that tasks already mastered are deemphasized and those requiring practice are performed. However, we do not know whether training managers systematically monitored and critiqued remedial training activities.

Finally, follow-up phone conversations revealed a complete lack of effort in reviewing and/or refining present evaluation procedures. The majority of field units in our sample did not send the U.S. Army Training Board a copy of their written evaluation reports. Nor did they complete the questionnaire which appears in the back of 71-2. To our knowledge, only one unit, Division B, assigned full-time staff with the responsibility of monitoring and refining their evaluation efforts, as indicated under IV.C.6.a. in Table 3-1.

3-3. SUMMARY; REFLECTIONS

This chapter reports results of use of a case study method to collect information as to how four divisions conduct battalion field evaluations. What we conceived to be mistakes and omissions were highlighted to better identify substantive areas to be covered in the Field Guide. This emphasis should not overshadow the fact that local initiative was evident as well—more exactly perhaps, well-springs of local initiative seeking further ARTEP related direction and clarification.

Let us, for the moment, consider these omissions and errors as possibly symptomatic of institutional problems. Institutions have policies, people and rules continuously in dynamic interaction. Within this are understandings as to what is policy, what is important, what are the proper ways to conduct business. Things get done because interlinked members of the institution share understandings as to policy, priorities and practices.

We suspect that to fully implement the changeover from ATTs to ARTEP some of today's understandings must be changed. And in these directions:

- Needed at all levels is a better appreciation of the complexity and difficulty of mounting sizeable field evaluations to derive best training value.

- This recognition should be reflected in emphasis on establishing *and checking out* E/C Group roles and procedures, and in more and better evaluator training. Evaluators learn as much, we believe, as the leader(s) of unit(s) being evaluated. Basically, however, the good evaluator is a multiplier. What people learn in the field depends largely on his careful observations, and how he provides feedback.

- Goals need to be established for improving field evaluations by team efforts between TRADOC, FORSCOM and units, and other responsible parties. In human learning, breakthroughs are rare; to be sought, rather, incremental gains and means of anchoring them. Here, it is necessary to develop measures of the extent to which goals are being attained.

These case studies serve as input to the next two volumes. Volume II analyses issues and problems surfaced and recommends solutions. A variety of scientific concepts can provide better insights into a number of these problems, suggest possible solutions, and means for evaluation of solutions. This material is not intended to be unit specific. It should, therefore, have broader applications than the battalion oriented Volume III Field Guide.