

AD-A274 753



(2)

IDA PAPER P-2881

QUANTIFYING THE MILITARY VALUE OF TRAINING
FOR SYSTEM AND FORCE ACQUISITION DECISIONS:
AN APPRECIATION OF THE STATE OF THE ART

Seymour J. Deitchman

October 1993

DTIC
ELECTE
JAN 07 1994
S E D

Approved for public release; distribution unlimited.

94-00476



INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under IDA's Central Research Program. Its publication does not imply endorsement by the Department of Defense or any other Government Agency, nor should the contents be construed as reflecting the official position of any Government Agency.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October 1993	3. REPORT TYPE AND DATES COVERED Final--September 1991--August 1993	
4. TITLE AND SUBTITLE Quantifying the Military Value of Training for System and Force Acquisition Decisions: An Appreciation of the State of the Art			5. FUNDING NUMBERS IDA Central Research Program	
6. AUTHOR(S) Seymour J. Deitchman				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 1801 N. Beauregard St. Alexandria, VA 22311-1772			8. PERFORMING ORGANIZATION REPORT NUMBER IDA Paper P-2881	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This paper reviews the state of knowledge that can enable military system and force planners to quantify the military contribution of collective training in military units (i.e., "unit training") designed to help the units succeed in battle. The purpose is to express the military value of hardware and of unit training in commensurate cost-effectiveness terms to ensure that the two are both considered and remain related to each other in system and force investment decisions. A review of the sparse relevant literature indicates that unit training under realistic conditions can increase key military capabilities of units from platoon to battalion size (and equivalents in the air forces) by factors of 2, on average. Hardware advances can increase military capability by like amounts if the requisite unit training is also provided. Without the investment in unit training the capability of a military unit should be discounted by about 50 percent. The training must be sustained continually to maintain system or force design capability. Since this will approximately double existing or new system life cycle cost over periods corresponding to a system life cycle, military budgeting and planning must be modified to account explicitly for unit training costs.				
14. SUBJECT TERMS collective training, cost effectiveness of training, military value of training, training, unit training			15. NUMBER OF PAGES 40	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR	

IDA PAPER P-2881

QUANTIFYING THE MILITARY VALUE OF TRAINING
FOR SYSTEM AND FORCE ACQUISITION DECISIONS:
AN APPRECIATION OF THE STATE OF THE ART

Seymour J. Deitchman

October 1993

Accession For	
NTIS	CRA&I ☒
DTIC	TAB ☐
Unannounced	☐
Justification _____	
By _____	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC QUALITY INSPECTED 8

Approved for public release; distribution unlimited.



INSTITUTE FOR DEFENSE ANALYSES

IDA Central Research Program

PREFACE

This report was prepared under the IDA Central Research Program. The author is indebted to Dr. Michael Fineberg, Dr. John D. Fletcher, Mr. Stanley Horowitz, Mr. Christopher Jehn, and Gen. Larry D. Welch for their most helpful review comments. Important words of appreciation are due to the late Dr. Earl Alluisi and to Dr. Jesse Orlansky for their suggestions and support that both urged and facilitated the author's work over the past 7 years that led to this paper.

ABSTRACT

This paper reviews the state of knowledge that can enable military system and force planners to quantify the military contribution of collective training in military units (i.e., "unit training") designed to help the units succeed in battle. The purpose is to express the military value of hardware and of unit training in commensurate cost-effectiveness terms to ensure that the two are both considered and remain related to each other in system and force investment decisions. A review of the sparse relevant literature indicates that unit training under realistic conditions can increase key military capabilities of units from platoon to battalion size (and equivalents in the air forces) by factors of 2, on average. Hardware advances can increase military capability by like amounts if the requisite unit training is also provided. Without the investment in unit training the capability of a military unit should be discounted by about 50 percent. The training must be sustained continually to maintain system or force design capability. Since this will approximately double existing or new system life cycle cost over periods corresponding to a system life cycle, military budgeting and planning must be modified to account explicitly for unit training costs.

CONTENTS

Preface	iii
Abstract	v
Tables	ix
Figure	ix
Acronyms and Abbreviations	xi
SUMMARY	S-1
I. INTRODUCTION.....	1
A. Purpose of the Paper.....	1
B. Need for Relevant Knowledge	2
C. State of the Capability	4
D. Topics Covered.....	5
II. EFFECTS OF VARIOUS FACTORS ON FORCE CAPABILITY	7
A. Time and Cost Patterns to Acquire Enhanced Capability	7
B. Intangibles	9
III. MEASURING THE MILITARY VALUE OF TRAINING.....	11
IV. SURROGATES FOR BATTLE	15
V. STATUS OF THE QUANTITATIVE EXPRESSION OF THE MILITARY VALUE OF UNIT TRAINING	19
A. NATO Defense Research Group (DRG) Symposium and Research Study Group (RSG) Reports	19
B. Military Value Work at IDA.....	20
C. Two Papers by Gen. Paul S. Gorman.....	23
D. ARI Analyses Using NTC Data	26
E. Additional Relevant Data	29
VI. COSTS OF TRAINING	33

VII. SIGNIFICANCE OF THE AVAILABLE "MILITARY VALUE" DATA.....	35
A. Force Improvement or Capability Reduction	35
1. Effectiveness Improvement or Decrement	35
2. Peak Effectiveness.....	36
3. Leadership: A Key Factor in Unit Effectiveness	37
B. Unit Training Cost and Interaction With Effectiveness Improvement	37
C. Impact of Unit Training on Force Planning: Broader Issues	40
VIII. INCORPORATING "MILITARY VALUE" CONSIDERATIONS IN SYSTEM AND FORCE ACQUISITION DECISIONS	43
REFERENCES	47

TABLES

1. Levels of Achievement	2
2. Comparison of the Properties of Three Unit Training Techniques.....	18
3. Implications of Available "Military Value" Data for Force Planning.....	41

FIGURE

1. Alternative Paths to Force Performance Improvement	7
---	---

ACRONYMS AND ABBREVIATIONS

ACEVAL	Air Combat Evaluation (an operational test program)
ACMI	air combat maneuvers instrumentation
AFB	Air Force Base
AH	Attack Helicopter
AIMVAL	Air Intercept Missile Evaluation (an operational test program)
ARI	Army Research Institute for the Behavioral and Social Sciences
ARTEP	Army Training and Evaluation Program
BG	Brigadier General
COEAs	Cost and Operational Effectiveness Analyses
DARPA	Defense Advanced Research Projects Agency
DIS	distributed interactive simulation
DoD	Department of Defense
DRG	Defense Research Group
FEBA	forward edge of the battle area
FX	field exercises
GPS	Global Positioning System
HQ	headquarters
IDA	Institute for Defense Analyses
LTC	Lieutenant Colonel
METL	Mission-Essential Task Lists
METT-T	Mission, Enemy, Troops (friendly), Terrain Control-Time
MILES	Multiple Integrated Laser Engagement System
NATO	North Atlantic Treaty Organization

NTC	National Training Center
O&M	operations and maintenance
O&S	operations and support
O/C	Observer/Controller
OPTEMPO	Tempo of Operation
RAND	a nonprofit national defense research organization
REALTRAIN	realistic training exercises for armored units instituted by Gen. Paul Gorman (USA, Ret.)
RED FLAG	Exercises on the Multi-threat Range at Nellis AFB, Las Vegas, Nevada, that employ TES and ACMI
RSG	Research Study Group
SIMNET	Simulation Networking
SOP	standard operating procedure
STRIKE U	Navy Strike Warfare Center
TES	Tactical Engagement Simulation
TOP GUN	Training employing TES and ACMI at the U.S. Navy Fighter Weapons School, Naval Air Station, Miramar, San Diego, California
TOW	Tracked, Optical, Wire-guided Missile (the 3000-meter range heavy Antitank Weapon introduced into U.S. forces in the early 1970s)
TRADOC	U.S. Army Training and Doctrine Command
USA	United States Army
USMC	United States Marine Corps

SUMMARY

Cost-effectiveness evaluations of military systems usually assume that design performance will be achieved. However, this achievement depends on the personnel who operate and maintain the systems. This paper summarizes the state of current, available knowledge that can enable military system and force planners to quantify the military contribution of the collective training of personnel in military units ("unit training" for short) designed to help the units win in battle. The paper addresses the problem of expressing the military value of hardware and of unit training in commensurate cost-effectiveness terms so that sound system and force investment decisions can be made by ensuring that the two areas of investment are considered and remain related to each other.

Full achievement of the desired application is still far in the future. However, fragments of understanding do exist. This paper integrates many of these fragments to yield some insight as to how to account for the hardware/training cost and effectiveness relationships. The methods permitted by the current state of knowledge are shown to be crude but serviceable.

First, three relevant introductory topics are discussed separately: effects of various factors on force capability; measuring the military value of unit training; and surrogates for battle. Available data applicable to the quantitative expression of the military value of unit training are then summarized. Some unit training cost issues are noted. The significance of the data and their application for the stated purpose are examined. Finally, the necessary research to make the methods more credible in the future is discussed.

The main results of this review are summarized briefly as follows.

A. EFFECTS OF VARIOUS FACTORS ON FORCE CAPABILITY

Added unit training effort can lead to force performance improvement in a short time (e.g., 6 months), but continuing expenditure for an indefinite period is necessary to sustain the new level of capability. The benefits of improving or replacing hardware in a force are not realized for some time after the acquisition decision has been made (e.g., 10-20 years); however, once gained, these benefits can be sustained with little additional expense *if associated unit training is maintained.*

Improved equipment, if it works as intended, will increase the proficiency of an entire force (if it is appropriately trained) by helping the force perform better in finding the enemy, by enabling it to reach the enemy more rapidly, and by making its attack more lethal. However, the equipment imparts only a portion of the improvement. Achieving these ends and winning the battle are also affected by many intangibles, including leadership, morale, motivation, ability to "read" the enemy and anticipate his actions, and cultural derivation of the force. What makes "aces" and "crack" units is understood intuitively by many senior military people but remains to be quantified.

B. MEASURING THE MILITARY VALUE OF UNIT TRAINING

Military system effectiveness evaluation is based on measures such as two-sided casualties, loss rates, kill rates, fractions of force surviving, battle line movement or territory occupied, forces captured, and so forth. Similar measures should be used to quantify and incorporate the military value of unit training in system and force evaluations. This may now become possible through the use of data from distributed interactive simulations (DIS) of combat and from instrumented exercise ranges.¹

To make both unit training and the measurement of its outputs manageable, the training and the outputs are considered by the training community in terms of tasks and subtasks that are completed in sequence to accomplish the overall activity. However, because of the intangibles sketched above, it does not necessarily follow that task proficiency will always lead to improved chances of success in battle. Therefore, most current measures of force performance in the field are based heavily on military judgment that is integrated in an orderly way for diverse tasks and quantified by expressing the judgments in terms of rating scales or scores. However, overall measures of success and the reasons for it might be lost in this process.

To quantify reliably the results from real or simulated combat with people "in the loop," two pivotal relationships must be examined in detail:

- The relationship between overall combat outcomes that would occur in warfare and the results that can be measured objectively
- The relationship between these measures of performance and judgments of performance by experienced practitioners.

¹ Computer simulations of warfare, known as "constructive simulations," are not included as sources of evaluative data for this purpose because they *require* data on force performance as input from outside rather than contributing it via the behavior of people manning the machines in the military force.

C. SURROGATES FOR BATTLE

Unit training takes place in operating time (running hours, flying hours, steaming hours), in DIS, and in field exercises. No single one of these surrogates for battle provides all the conditions and verisimilitude that are necessary for teaching units how to behave in battle and for measuring success. Only instrumented field exercises and DIS exercises with fully manned units in situations of free play can suggest the outcomes of battles as they are affected by variability of behavior and the uncertainties people impose on each other. However, neither mode is complete in itself. Field exercises convey the "feel" of the equipment in the environment (dust, vibration, odor, noise, breakdowns); DIS allows easy expansion of virtual battle space and the ability to manipulate experiments inexpensively; and both allow scope for command, control, and maneuver of units of various size. Operating time contributes to proficiency with hardware and unit operating skills needed for DIS and field exercises. Hybrid simulations involving both DIS and field exercises can provide elements of both for greater realism in some aspects of training. ("Pure" computer simulations, now known as "constructive simulations," are used as training devices in some circumstances, but they are excluded in this paper because they are considered to be tools for evaluation by calculation rather than man-in-the-loop training devices that can be used as surrogates for battle in unit training.)

D. STATUS OF QUANTITATIVE EXPRESSION OF MILITARY VALUE OF UNIT TRAINING

The data and analytical results reviewed include the following:

- Reports from a North Atlantic Treaty Organization (NATO) symposium and a Research Study Group (RSG) on the military value of training
- Two studies performed at the Institute for Defense Analyses (IDA), which include and integrate extensive data available from earlier relevant work, including the NATO reports
- Two papers on the topic by Gen. Paul S. Gorman (USA, Ret.)
- Army Research Institute (ARI) analyses of data from the National Training Center (NTC)
- A RAND evaluation of mission training in the AH-64 attack helicopter
- A U.S. Army comparative evaluation of training effectiveness in DIS and field exercises, using tank and mechanized infantry platoons.

E. COSTS OF TRAINING

The most frequently, accurately, and directly measured cost of training encompasses classroom, simulator, flight, and professional development costs for individual skill training. These are the costs included in training budget listings. Costs of training during routine running hours, flying hours, or steaming hours are ordinarily embedded within Operations and Maintenance (O&M) budgets. Pay and allowance costs of military personnel, whether in training status or not, are included in military personnel cost categories. Costs of unit training in DIS and field exercises tend to be included in non-training budget categories, such as those for facilities and forces. Association of field and DIS unit training costs with systems in new-system Cost and Operational Effectiveness Analyses (COEAs) is usually derived from planning factors rather than from a specific, prospective analysis of the new systems in military settings. The planning factors reflect prior experience rather than prospective need. This practice derives from a desire to gain experience with new systems before projecting unit training needs. To put the system-related and training cost algorithms on a comparable basis for all of the Service COEAs that precede system acquisition and force planning decisions, military budgeting methods will have to be rearranged so that the projected and actual costs of unit training can be determined uniformly and reliably.

F. SIGNIFICANCE OF THE AVAILABLE MILITARY VALUE DATA

A review of the available data leads to the following insights toward quantifying the military value of unit training:

Effectiveness Improvement

- Realistic training [what Gen. Gorman has called Tactical Engagement Simulation (TES), which is based on mock combat against an intelligent enemy in real or simulated terrain] at levels from platoon to battalion can improve ground force combat capability, on average, by factors of 2.
- Realistic training consisting of squadron level tactical bombing practice can increase tactical bombing accuracy with particular aircraft by about a factor of 2 when flight hours to practice are increased from about 10 to 40 hours per month.

In military combat areas essential to battlefield success, equipment replacement can also achieve factors of 2 or better improvements in force capability. Unit training effects add to these improvements. Conversely, it can be considered that:

- If there will not be adequate unit training to achieve and maintain peak unit combat skills (i.e., the "design" levels of performance) with new equipment, then the performance of the equipment should be discounted by a factor of 2 in assessing a new system's cost-effectiveness. This is a rough approximation in the absence of more precise estimates applicable to equipment of particular kinds and missions.

Peak Effectiveness

- The asymptotic value (i.e., the value where further training yields no returns) of peak training starting from some low sustained level in armored combat may be reached after four or five exercises in a training session. It may be reached at three training sessions per month in helicopter anti-armor combat. As a matter of judgment, asymptotic training effectiveness is said to be retained for a few months. Very roughly, the available data on asymptotic values of training effectiveness suggest the need for a realistic unit training cycle at least once per year, but the data are insufficient to confirm this or some greater frequency as the appropriate one.
- Asymptotic values of tactical bombing effectiveness appear to depend at least as much on career flying hours in a specialty (more than about 1500 hours) as on immediate practice, becoming nearly twice as good at that stage as accuracy at the beginning of the career specialty--e.g., at 300 hours. The additional factor of 2 gain with immediate practice (in going from 10 to 40 flying hours per month) raises the asymptote; that is, it adds to the capability derived from the career level of achievement, so that highly experienced pilots with recent extensive practice can achieve up to 4 times better accuracy, on average, than inexperienced pilots who have not had recent practice.
- As overseas force deployments decline, the impact of reducing crew turbulence (moving and rotation) on retention of unit effectiveness should be a subject for exploration and tracking.

Leadership: A Key Factor in Unit Effectiveness

- It appears that the qualities of good leadership can be defined from data emerging from TES and can be determined for individuals by adapting methods that have been developed for personnel selection and assignment. The definition should be refined, specific tests should be devised, and appropriate assignment practices should be undertaken. Currently, the leadership selection process is informal. There would be value in formalizing the process to increase the success rate in selecting and assigning potentially high-quality leaders entering the military in times of reduced force levels.

Unit Training Costs

- The available analyses show that unit training costs to achieve and sustain the "canonical" factors of 2 force effectiveness improvements are, coincidentally, roughly equal to system life cycle costs over, say, 20 years. It is generally accepted that life cycle cost, including operations and support (O&S), will be about 1 1/2 to 2 1/2 times as high as hardware acquisition cost alone. This result is also coincidental, but it is often used as a rough rule of thumb when new system planning is initiated. The magnitude of sustained unit training cost implies another rough rule of thumb: that system life cycle cost estimates must be increased by about another factor of 2 to determine the total cost of keeping a system and its personnel in the force at peak capability. If funds for sustaining levels of unit training are not budgeted, then force capability *with either new or existing systems* will not be maintained, within roughly a factor of 2.
- Two kinds of hardware and unit training costs—equipment upgrades and unit training in DIS—are considerably cheaper than full equipment replacement or field training. Specific-system analyses and further unit training research are necessary to ascertain the full extent of substitutability of the less for the more expensive approaches.
- Systems being replaced may have been subject to unit training at the required level, with assignment of attending budgets. However, new systems may require different amounts of unit training, with changes in associated unit training costs. Costs such as major-force retraining for new tactics and doctrine also must be included. Discounting new system performance for lack of adequate unit training budget should, therefore, be done judiciously to account for the history of the existing system and the anticipated demands of the new system. Added unit training costs could be needed for either the current or the new system if unit training with the existing system was not at appropriate levels.

G . APPLICATION OF QUANTITATIVE "MILITARY VALUE" DATA

To be more precise in planning, there must be more understanding of the relationship between training input and performance output. To help achieve this, two areas of research must be pursued:

- The validity of field exercises and of DIS training in representing the results of combat operations and the contribution of each to unit training must be established.
- The costs of unit training in TES, for both field exercises and DIS, must be aggregated and considered as budget categories that can be managed.

Apart from the obvious possibility of force expansion, which entails increased budgets, extensive time, and a degree of political turmoil, changing budget allocations within a given overall budget can allow extensive force planning versatility. If the costs were known, it would be possible to allocate unit training funds to maintain one of the following:

- A relatively small but highly ready force able to go into combat on short notice
- A large force requiring additional (e.g., "topping off") training to go into combat
- A force with some units, in numbers allocated by plan, in each status.

Planning for forces to be able to undertake effective, low-casualty combat operations on short notice is essential in the new strategic environment for national security, and it cannot be accomplished without taking such training costs into account. Giving visibility to the unit training budget would give an element of insurance against allowing a force to go "hollow" through the budgeting and funding allocation process.

I. INTRODUCTION

A. PURPOSE OF THE PAPER

Techniques to quantify cost-effectiveness relationships for military equipment, systems, and forces are known, although the methods—even after 50 years of work—have many shortcomings and needed data are still absent. Less is known about how to quantify such relationships when human factors and training are involved. This paper summarizes the current, available knowledge that can enable military system and force planners and analysts to quantify the military value of unit training and indicates what must be done to assign such a value more credibly than is now possible. The purposes of placing a military value on unit training are as follows:

- To be able to estimate the amount and costs of unit training necessary to help forces succeed in battles and campaigns, so as
- To be able to account for human capability in using military systems when making cost-effectiveness evaluations of the military systems and the forces that use them, and thereby
- To allocate adequate budgets and funds to achieve the stated force objectives.

A period of intensified applied research on the selection, training, utilization, and performance of military personnel began during World War II. Until about two decades ago, this research focused mainly on individual training and performance that could be measured in the laboratory. Since that time, there has been increasing recognition of the importance of collective training in units (or as it will be called in this paper, unit training). However, since unit training has been implemented in the field rather than in more easily controlled environments, most current understanding of the impact of such training is based, in the last analysis, on anecdotal evidence and military judgment. To understand the nature of the phenomena that underlie training effects military judgment is very important, but the magnitudes and directions of those effects can only be described accurately from the quantitative measurement of crew and force behavior in real or simulated operations. The quantification of input and output for unit training that leads to the quantitative analysis of group performance has become possible only very recently through DIS and through instrumented field exercises.

Consider a hierarchy of activity associated with military conflict, as depicted in Table 1.

Table 1. Levels of Achievement

Military Mission	Example Task Within the Mission	Example Measure of Performance
1. Individual or crew engage the enemy	Shoot at and hit a target	Number of hits
2. Operate a military unit	Move a unit from A to B	Adherence to plan
3. Succeed in battle	Defeat or delay enemy force	Casualties inflicted and received; territory occupied; delay achieved
4. Win a war	Destroy enemy army; occupy enemy country	Objectives achieved or not

Each mission level in this table encompasses the one that precedes it. The table establishes a convenient shorthand for describing the hierarchy of force performance under discussion in this paper, and the task or mission levels it defines will be used throughout.

Even to reach the stage of arguing about the achievement of ultimate objectives, military or political commanders must have forces that can succeed in the narrower military parts of their missions. Thus, a question remains: How does the military invest to accomplish this? Many kinds of investment are needed; however, for purposes of this paper, they can be described as investments in hardware, in training, or in all other military activities. This paper addresses the problem of expressing the military value of hardware and of unit training in commensurate cost-effectiveness terms so that sound system and force investment decisions can be made by ensuring that the two areas of investment are considered and remain related to each other.

B. NEED FOR RELEVANT KNOWLEDGE

Much is known about how to quantify the effects of training when evaluating individual and job proficiency and when selecting people to perform certain tasks, but quantitative descriptions of the relationships between unit training and success on the battlefield are scarce. In the past, quantitative research on unit training has tended to focus heavily on specialized instruction and practice to improve the performance of military tasks

at the first two levels in Table 1, because such performance can be described with some precision and the degree of successful task completion can be measured and rated. However, recent work by the ARI, Monterey, is beginning to shed some light on how training activity at the first two levels relates to an approximation of the third level, i.e., success in battle as determined in mock combat on the instrumented ranges of the NTC.

Currently, system acquisition decisions are based on the assumption that equipment will work as designed once it has been integrated into the forces, but this can be true only if there are appropriate training expenditures and methods. Comparisons of alternative equipments are based on the hardware costs only and rely on design performance projections to examine whether an investment in new equipment should be made or which choice among several available equipments might make the best investment. System budgets are allocated specifically to tactical air squadrons, ships, or tank battalions. For life cycle costing, all systems carry the costs of operations, maintenance, and integrated logistic plans. These costs involve mainly the fabrication, transportation, and storage of hardware (spares). Projected life cycle costs used in evaluating a new systems' potential worth also include allocations for proficiency training at the individual and crew levels. For example, this will include costs of system simulators, which are essential for training with most advanced systems. In system acquisition and life cycle costing, training is budgeted for new classroom training and for in-unit running hours for land forces, flying hours for air forces, and steaming time for ships.

Unlike costs of hardware, which can be estimated for any new system (with uncertainties affirmed by continual management conflict about cost overruns), proposed training costs tend to be assembled largely from precedent, i.e., expenditures in prior budgets and modest changes to them, and tend to reflect planning factors emerging from military judgment and base support considerations in which precedent plays a large part. For example, flying hours for a new fighter aircraft system are based on the characteristic flying hours allocated to fighter squadrons in previous and current defense budgets rather than on an estimated number of hours necessary to achieve full combat proficiency with the particular new system.

In addition to the components that are usually tracked, the full training costs for new systems that have not been wrung out in the field would normally require the inclusion of special training to make units battle ready, such as periods at the NTC, RED FLAG, TOP GUN, the Navy Strike Warfare Center (STRIKE U), and the United States Marine Corps (USMC) Air-Ground Warfare Center, and training periods at sea using embedded

simulation modes for shipboard systems. These costs are not explicitly accounted for in connection with the new systems but are, rather, embedded in different parts of the defense budget. Since it is usually essential to incur these costs for a new system to achieve the performance credited to it in the initial system evaluations, masking the training costs in the defense budget means they are difficult or impossible to account for and to track both in system acquisition and in force planning decisions.

The hardware part of the budget is determined by quantitative methods that have some basis in engineering and analysis. There are many shortcomings in this part of the budget process, and these shortcomings are known to the sophisticated practitioners, who (management politics permitting) allow for them almost intuitively in application. However, the human part of the budget—an amount approximately equal to that of the hardware acquisition part of the overall defense budget—is based on what might be called educated guesswork. Since the total budget for a system or force remains incomplete until the cost of the human element has been incorporated, establishing quantitative relationships between unit training and unit performance output is essential for both aspects of the budget—the hardware part and the human part—to be considered on at least a roughly equivalent basis. In turn, this would result in more informed investment decisions and would lessen the vulnerability of training budgets to arbitrary reductions in times of budget stress.

C. STATE OF THE CAPABILITY

Unfortunately, the desired capability to quantify training budgets in terms and at levels of precision commensurate with hardware budgets does not exist. Because the effort needed to attain this capability would be roughly one or two orders of magnitude larger than the levels of effort typically allocated for training research, the knowledge probably will have to be gleaned as a by-product of training activities undertaken for operational reasons, with associated constraints on the data obtained and the availability of the data for such applications. Therefore, it can be anticipated that full achievement of the desired capability is still far in the future.¹

¹ For example, the report on results from the NTC (Keesling et al., 1992), was 5 years in the making. Its publication was delayed because of the time taken to build a large enough sample of units trained and because of constraints imposed by the Army on use of the data from training exercises to ensure that units could not be identified for specific unit or commander evaluation purposes. Also, Hammon and Horowitz (1987) have shown that extensive performance databases are maintained by the Services but not with centralized access for exploitation in the current context. Although such databases could be explored for the purpose, part of the constraint on using them is similar to that described previously for the Army.

Fragments of understanding do exist. This paper attempts to summarize and integrate these fragments to provide some insight about the hardware/training cost and effectiveness relationship. The methods permitted by the current state of knowledge will be shown to be crude but serviceable, given the kinds of approximations that are often necessary for "paper systems" acquisition and judgment-based force planning activities.

D. TOPICS COVERED

This paper initially reviews three relevant introductory topics:

- Effects of various factors on force capability
- Measuring the military value of unit training
- Surrogates for battle.

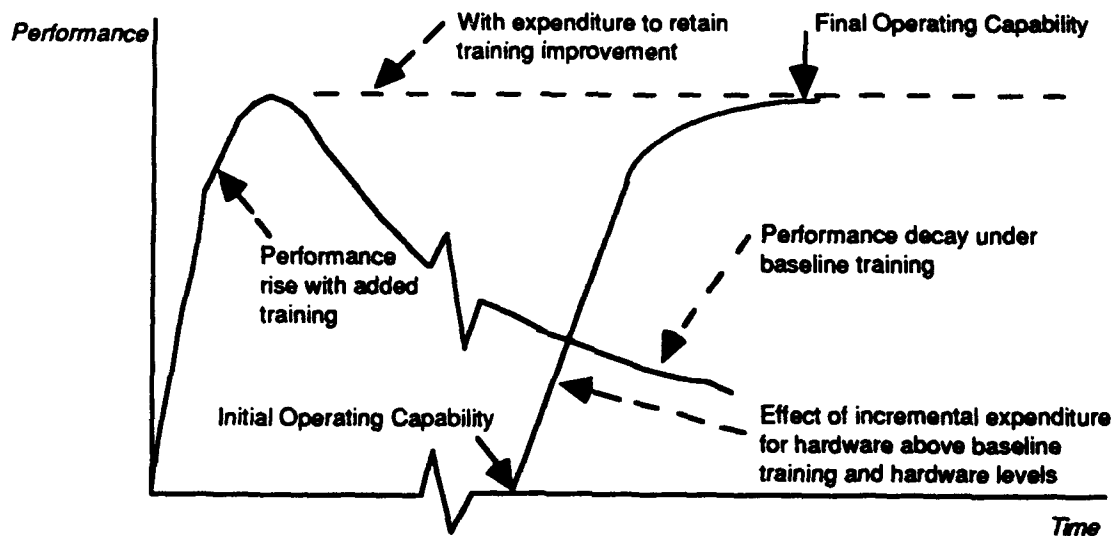
The paper then reviews the applicable data that are available to quantify the expression of the military value of training and its application to system acquisition and force effectiveness decisions. Finally, the significance of the data for the stated purpose, work by the Services that is necessary to improve the state of the data and the state of the art, and approaches to applying the data are reviewed.

Not surprisingly, an attempt to quantify the military value of unit training proved to be a rather amorphous topic. In the attempt to focus on the effectiveness and cost aspects of unit training, other interesting aspects of this topic, which are separate from the effectiveness and cost issues but often bear on these issues in some way, were uncovered. Where appropriate, these digressions are indicated and presented as illuminating aspects of the training problem that may have some relevance in the broader picture of application in the real world.

II. EFFECTS OF VARIOUS FACTORS ON FORCE CAPABILITY

A. TIME AND COST PATTERNS TO ACQUIRE ENHANCED CAPABILITY

It is obvious that new and improved equipment will increase the capability of a force (if the equipment acquisition is wisely thought through) and that people will have to be trained to use it. It is not always obvious that the equipment and training contribute differently to the level and timing of force proficiency improvement. Assuming constant baseline training expenditures, added training effort can lead to force performance improvement in a short time, but continued expenditures for an indefinite period are necessary to sustain this new level of capability. On the other hand, the benefits of improving or replacing hardware in the force are not realized for some time after the decision has been made; however, once gained, these benefits can be sustained with little additional expense if the associated baseline training level is maintained. The contrast is illustrated in Figure 1.



Note: Performance improvements shown equal;
no implications about relative expenditures for training or hardware

Figure 1. Alternative Paths to Force Performance Improvement

Under today's conditions, equipment takes years—at least 10 and as many as 25—to be acquired and introduced into the forces. There is time to prepare training and logistics programs and to plan cost profiles. The training programs thus planned are designed to teach individuals and crews the mechanics of using the new equipment. However, in the early stages of equipment planning, there is usually little understanding about the amount of unit training that will be necessary to ensure that the equipment will be used to maximum effect during combat. Moreover, new system acquisitions often transform the configuration of the force, and this implies further costs to effect the new tactics and doctrine that must be documented, practiced, and learned (i.e., trained) throughout the military forces. In addition to the costs incurred, these changes require an expenditure of time for the forces to become proficient in the new modes of operation.

There is much anecdotal evidence—the most recent emerging from Desert Shield and Desert Storm in 1990 and 1991 (United States Department of Defense, 1991)—to indicate that intensive unit training over a short period of time (for example, 6 months prior to Desert Storm) can increase force capability dramatically.² However, to retain the capability, the training must be repeated periodically. Currently, the asymptotic values of training and cost to achieve the level of unit training that maximizes unit capability without expending resources for no returns, and the periodicity of training required for units to retain proficiency, have not been determined for many military applications. There is a better understanding of the cost for achieving individual and unit task skills (levels 1 and 2 of the hierarchy in Table 1) than there is for unit mission skills (level 3 of the hierarchy). Since budget planning factors obviously cannot be available when the required time and effort to achieve a specified level of capability are not known, such factors, if used, are mainly reflections of historical expenditure levels. However, these levels will be distorted by expedient and usually arbitrary changes in training expenditures over the years. Thus, when a surge in training must take place, the funding is usually found *ad hoc* but the pressure to sustain the training relaxes when the need disappears. This is not a healthy position for U.S. armed forces in a period when they may have to respond very rapidly to crises, often without the opportunity to engage in the unit training that occurred before the Desert Shield/Desert Storm operation.

² Also, discussion of training an artillery unit on the way to the North African campaign, Gen. Donald Bennett, USA (Ret), private communication to the author.

B. INTANGIBLES

Observers tend to judge potential unit capability in battle by a unit's success in accomplishing contributory military tasks for which proficiency can be measured. Quantifying the value of unit training at the third level of the hierarchy in Table 1 and assessing the effectiveness of a certain amount of expenditure on unit training is difficult because the military art involves much more than simply gaining proficiency with equipment and learning how to drill with this equipment in unit formations (level 2). Such skills are internalized in any environment, military or civilian. As a simple and obvious example, driving an automobile is essentially the same whether the vehicle is military or civilian. Performance using the automobile as a tool is related to mission success—e.g., reaching a specified destination at a specified time—in which more or less successful task accomplishment—route planning, map and clock reading, driving—is implicit. Intangible factors that extend task skills into mission success determine how well a force will move from success at level 2 to success at level 3.

Improved equipment, if it works as intended, will increase the proficiency of an entire force by helping the force perform better in finding the enemy, by enabling it to reach the enemy more rapidly, and by making its attack more lethal. However, the equipment imparts only a portion of the improvement. Achieving these ends and winning the battle are also affected by many intangibles, including leadership, morale, and motivation. There are many instances in history when forces having advanced equipment did not perform as well as forces with less advanced equipment because they were deficient in the other qualities. The early Arab-Israeli wars furnish examples. (See, for example, Dayan, 1967)

In addition to these intangibles, units must learn to "read" the enemy to know how to attack him successfully by exploiting their strengths and the enemy's weaknesses, how to effect surprise, how to survive, and how to interact with each other in such a way that force can be applied in the right amount, at the right time, and in the right place. These skills must be developed during unit training in realistic (actual or simulated) environments against "live" enemies to improve combat performance of the force being trained.

It has also been observed that in military forces a few individuals—"aces"—manage to achieve the greatest number of kills against the enemy whether it be in air warfare, tank warfare, or infantry close combat. It is also known that some units, from divisions to individual ships and submarines, achieve better success in battle than other units. What makes a "crack" unit is understood intuitively by many senior military people but remains

to be quantified. Clearly, it is a function of unit training, a synthesis of skills by individuals learning how to work together, and the synergistic effects of leadership, discipline, and practice. Quantification of the individual and group qualities that lead to the development of crack ground force units may emerge from ongoing work being conducted by the ARI using data from the NTC. Analyses of data from other Service training centers may provide insights for air and sea forces. The knowledge gained from such analyses will still have to be related to investment in training to know how to bring crack units to the fore and how to create more of them.

As a sidelight with implications for training-hardware tradeoffs, there are indications that military equipment that substitutes automatic functions for manual tasks reduces the time for the skills required to reach expert levels. Tank gunnery tests of individuals in different score categories on the Armed Forces Qualification Tests showed that Category 4 (low-scoring) troops required much less training to reach the same proficiency levels as Category 1 (high-scoring) troops in the more automated M1 tank than they did in the M60 tank (United States Military Academy, 1984). Data also indicate that pilots with a few hundred hours of mission experience using fully automatic fire control and bombing systems attain bombing accuracies comparable to those achieved using manual modes by pilots with thousands of hours of experience (Hammon and Horowitz, 1992).

Modern equipment is designed with more flexibility and versatility because advancing technology makes it possible to embed routine but attention-demanding tasks, such as aim-point tracking and bomb release, in automatic modes. Once the automatic modes are exploited for early achievement of operational skills, emphasis in training time and cost budgets can be shifted from individual skill training to honing other skills that are important to unit performance in combat, such as surviving, tactics, integrating accurate fire with maneuver, intra- and inter-unit coordination, and exercise of command and control. Thus, within given budgets, making equipment easier to use can advance the less tangible skills that lead to successful combat outcomes and that require enhanced unit training to develop.

III. MEASURING THE MILITARY VALUE OF UNIT TRAINING

Measures from computer models of force interactions, such as TACWAR and JANUS, that are used to evaluate military systems include quantities such as two-sided casualties, loss rates, kill rates, fractions of force surviving, movement of the forward edge of the battle area (FEBA) or territory occupied, forces captured, and so forth. If the military value of unit training is to be quantified by methods similar to those used for assessing hardware cost effectiveness, then similar measures should be used. Through the use of the data from distributed interactive simulations of combat [e.g., Simulation Networking (SIMNET³)] and the data from instrumented ranges, like the one at the NTC and other ranges, these measures are now available to describe ground force combat. Analogous position measurement systems for air exercises, with the calculation of terminal engagement outcomes, were developed in the 1970's in connection with the AIMVAL and ACEVAL air combat maneuvering tests, and they are used at RED FLAG and similar activities of the other Services. Measures of bombing and gunnery accuracy are routine and can also be applied to assessments of unit combat outcome in air-to-air and air-to-ground warfare through application to engagement simulations. Ship performance during training exercises is rated by judgment-based scoring systems and the records are maintained (Hammon and Horowitz, 1987). However, these data may not be as concrete as those of the ground and air units.

The ground forces' essential skills are embodied in the triad "move, shoot, and communicate." Ground force units must integrate their ability to maneuver in coordination with adjacent units, to use weaponry effectively in coordinated fire to destroy opposing targets that result from the maneuver, and to communicate effectively to ensure the coordination. Combat skills for the air forces would include maneuvering to engage the enemy; using gunnery or missiles against enemy targets in the air; using guns, unguided bombs, or guided weapons to shoot effectively at enemy targets on the ground; and using communications for coordination and mission planning. Naval combat skills involve the

³ SIMNET is a DIS network involving tank, other armored vehicle and close air support aircraft simulators. (See Alluisi, 1991)

operation of intricate machinery (i.e., ships and shipboard equipment) and the exercise of those skills necessary for airmen and ground forces (Marines) to succeed in combat. For all forces, command and control uses understanding in all of the areas listed above to exercise effective management and control of force operations in warfare.

Many of the skills involved in these activities are individual skills. In addition, the operation of major platforms, such as tanks, aircraft, and ships with their intricate weapon and communication systems requires team activity and the effective training of teams to perform these activities. Operations as diverse as maneuvering a tank unit into marching order on a road and deploying it properly for battle at the appropriate time and place or launching and recovering deckloads of airplanes from aircraft carriers are highly complex in terms of the overall skill level requirements of individual crewmembers and their leaders at different echelons of command. To make the manning and training operations manageable, they are divided into tasks and subtasks that, taken in sequence and together, add up to accomplish the overall activity. For example, the breakdown of military activity into tasks is a key element of the Army Training and Evaluation Program (ARTEP).

Because of the intangibles sketched in the previous section, proficiency in tasks will not always lead to improved success in battle. Different countries perform unit training in different ways, but some countries win more often than others. This might be because their training systems for missions and tasks are better, or there may be other determining factors. In an analysis of the level of casualties needed to put a unit out of action, Wainstein (1986) demonstrated that there are marked national differences in force capability, persistence, and the ability to succeed that seem to transcend both the amount and kind of training that these units have had and their experience in combat. The oft-mentioned but ill-defined intangibles of leadership, motivation, and morale, and one not often mentioned, culture, clearly play important roles. But, if task performance is not a defining variable, how, then, can one measure the ability to win?

Most current measures of force performance in the field are based heavily on exercise of military judgment, integrated in an orderly way for diverse tasks and quantified by expressing the judgments in terms of stratified scales or scores. Examples include the Mission, Enemy, Troops (friendly), Terrain control-Time (METT-T) scores used at NTC (McFann, 1990); Observer/Controller (O/C) unit performance ratings at NTC (Keesling et al., 1992); judgment of skill levels in carrier landings (Horowitz, Hammon, and Palmer, 1987); and Gorman's evaluation function, $R = f(W,P,T)$, where R is readiness (taken to be combat effectiveness), W is weapon effectiveness, which can be measured, P is individual

or unit proficiency, which can be measured, and T is "technique," which must be assessed by judgment (Gorman, 1990). The differences in performance that lead to "aces" in every area, from air-to-air combat to shipping kills by submarines, are also measured by judgment, which can be an uncertain guide. Gorman (1990) describes how battalions that he expected to do well did not, and vice versa, in Division field training exercises. He attributes this outcome to unexpected qualities, positive and negative, of the leadership and to unanticipated effectiveness of the units' prior training.

A careful reading of the literature suggests that broad military judgments seek first to express the degree of successful mission completion. Then, the components of success or failure are examined to assess what led to the outcome. In a different application, judgment-based rating systems that express performance on a rating scale tend to relate to the mission component level. This has explanatory value that might be lost in overall rating schemes. For example, anomalies such as mission success when intermediate steps were not done well, or mission failure when they were, would not be explained by broad judgments of overall performance. However, when components alone are rated, overall measures of success might be lost.⁴

Objective measures of overall success or failure, such as occupying terrain or capturing the enemy, assume paramount importance only in wartime, but they may be available from exercises at ranges such as the NTC or from DIS. Effects of joint combined arms combat, such as combat by ground forces having, or not having, close air support, might also be determined from such ranges and simulations, if they were sought.

During realistic training of units and commanders, a distinction must be made between using quantitative measures to evaluate equipment and training effectiveness and using judgment to assess training effectiveness. Quantitative evaluation and judgment are expected to be helpful in both cases, but their relative positions are reversed. Quantitative measures are essential for assessing the relative value of investments to avoid arguments based on opinion when billions of dollars may be at stake. However, the numbers must be tempered by judgments about the extent and conditions of their applicability. On the other hand, judgment by experienced practitioners is the only way to capture the intangibles of

⁴ See, e.g., Keesling et al., (1992). This report compares overall success rates at NTC, as judged by O/Cs, with thoroughness in planning and performing tasks specified by unit leaders according to principles of training prescribed in Army Training Manual FM 100-25. There is only partial correlation between overall outcome in engagements and how closely the units followed prescribed tasks and procedures. For additional discussion of this point, see Section V-D of this paper.

leadership, motivation, morale and the personality of the commander in advancing skills through training. However, quantification of outcomes can help temper these judgments by introducing a measure of objectivity.

To quantify reliably the Table 1 "third level" results from real or simulated combat with people "in the loop," two associated relationships remain to be established:

- The relationship between overall combat outcomes that would occur in warfare—terrain occupied, troops captured, forces destroyed—and the results that can be measured objectively—losses, kills, times of events, numbers of communications
- The relationship between the measures of performance that can be obtained by range or simulator instrumentation—losses, kills, and so forth—and judgments of performance by experienced practitioners.

Recent results from the NTC⁵ indicate that measurable events and combat outcomes are correlated, but the strength and extent of applicability of the correlation have not been fully established. Wainstein's results suggest that establishing the correlation firmly may not be easy because forces from different countries with different experience react differently to casualties when planning and implementing their future activities. Similarly, a historical review by Gorman (1992), aimed at illustrating the intangibles of training, suggests that combat experience and changes of leadership can affect the overall outcome of a battle when objective performance factors such as kill ratios do not change very much.

Both range exercises and DIS can help to establish the relationships between detailed performance measures in DIS and in the field and between observed combat outcomes and judgments about combat outcomes in either medium. This, in turn, would lend confidence that the factors derived to express the effectiveness of unit training, whether changed hardware is involved or not, could be used with some validity in computer simulations (constructive simulations) of warfare. Constructive simulations are the preferred instrument to evaluate system and force changes over a wide range of characteristics and scenario parameters because they can be manipulated so easily and inexpensively.

⁵ Discussed in Section VII-B of this paper.

IV. SURROGATES FOR BATTLE

While actual engagement in warfare indicates the effectiveness of unit training, along with the effects of many accidents of circumstance and "luck," good and bad, it does not offer the opportunity for rigorous data gathering and analysis to show how to express the results of the training or how to make the training more effective. This must be done in some simulation of battle, which acts as a surrogate and permits the time and expense of measurement and replication. Essentially, the surrogates for battle exist in three areas:

- Operations in a base area setting, or, for the Navy, in transit to operating areas, known in the Army as OPTEMPO and in the Air Force and the Navy as Flying Hour and Steaming Hour Programs
- Exercises away from base areas (including field exercises, command post exercises, and games)
- Simulation with people "in the loop," using simulators and simulator networks (DIS) for unit and individual training.

["Constructive simulations" also can be considered as devices for some kinds of training of military commanders. However, for purposes of this paper, the constructive simulations are considered decision-assisting tools in which the people are external to the simulation. The surrogates for battle of concern here are the simulations in which the people are embedded, as they are in military units and in large-scale military systems. In command and control system designs, the trend is to incorporate constructive simulations into high-level military headquarters as decision aids, and the use, simulation, or replication of these headquarters would incorporate the constructive simulations, which would become part of the larger systems being operated or simulated in the field or in DIS.]

No single one of the surrogates for battle listed above gives all the conditions and verisimilitude that are necessary for teaching units how to behave in battle and for measuring wins or losses. Base-area operations (OPTEMPO, flying hours, and steaming hours) do not necessarily train units for combat, but they do provide training in the operation and maintenance of equipment, coordinated movement and maneuver, unit cohesion, and related matters. The second and third categories of surrogates listed above allow commanders to work with their staffs and with each other to learn and to practice the

"drill" of planning the activities of their units in coordination with friendly activity and in the face of a substitute enemy (e.g., another unit or a referee's input). Only field exercises and DIS or hybrid exercises involving a combination of field operations, which may or may not be instrumented, and DIS can allow commanders at levels up to Divisions, Corps, or Armies (with analogous echelons in the other Services) to work with some or all of their troops at all echelons below the highest one represented. Whether this is done depends on the resources available for the exercises.

Manned simulators used for individual proficiency training must be accurate (and therefore often elaborate) replications of the actual equipment so that the individuals can learn how to use that equipment and to be as proficient with the simulator as they would be with the real equipment. For unit training, however, networked simulators such as the SIMNET or actual weapon systems with embedded simulation modes enable individuals to work together to accomplish things that they can accomplish only in units. In such cases, the fidelity of the individual equipment is subordinate to the fidelity of situations.

A comparison of instrumented field exercises, DIS, and hybrid exercises involving both is of special interest in the context of this paper. Only these techniques allow the exercise of fully manned units in situations of free play that can suggest the outcomes of battles as affected by the variability of behavior and the uncertainties that people impose on each other. Because the physical environment of the field exercises approximates that of actual combat more closely, the tendency is to assume that the field situation is closer to actual combat than the DIS situation. This assumption may or may not be true, but the question—What resembles actual combat more closely: a field exercise or DIS?—does not have to be phrased in this manner. From the unit training point of view, the comparison should be based on the situation and force characteristics that can be reproduced in each approach and on the aspects of training to which those characteristics contribute.

A hybrid exercise might be especially useful for exercises involving large units if the ranges are not available but the functioning of higher headquarters and their overall command and control abilities need to be examined or practiced. This exercise would involve some units in the field so that the physical stress of the environment—time delays and deviations imposed on the plan by the environment—affect the units' assembly and operations. The exercise would also involve some manned DIS units, which would contribute to the uncertainties caused by human variability and would increase the headquarters' load by requiring them to manage the movements, communications, and planning of both kinds of units (those fielded and those in the DIS). To a large extent, the

entire hybrid operation can replicate many of the conditions and stresses of very large maneuvers, with fewer of the resources and range constraints that would characterize such maneuvers if they were held as pure field exercises.

Table 2 compares the three approaches.

Table 2. Comparison of the Properties of Three Unit Training Techniques

Quality	Field Exercises (FX)	DIS	FX/DIS Hybrid
Combat Stress	• No fear of being killed, wounded, captured • Stress of competition, with physical environment stresses imposed (see below)	• No fear of being killed, wounded, captured • Stress of competition; physical environment stress much less (see below)	• No fear of being killed, wounded, captured • Competition is mix of exercises and DIS • Competition of the battle felt in higher headquarters
Physical Environment	• Heat, cold, dust, vibration, glare, odors, physical exertion • Deterioration of the machinery • Physical, electronic, chemical noise • Maneuver space and cost limit unit size	• Not accurately reproduced for fielded units; can be accurate for headquarters units • Can add stressors to simulators to approximate conditions, depending on simulator fidelity • Numbers of simulators (a cost item) limit unit size	• Some units exposed to and operate in physical environment • Some units operating in DIS mode • Environmental effects on each type unit correspond to type of simulation; may be some interaction; effects unknown
Equipment Fidelity; Crew and Individual Proficiency	• Accurately reproduced; can include live fire, in a firing range mode • Does not include incoming direct or indirect fire	• Approximated, unless accurate simulators are used; latter could be embedded simulator mode of actual system, (esp., shipboard), but verisimilitude usually given up for numbers	• Reproduced as in DIS or as in field exercises for the units in each type simulation • Can accurately reproduce operating conditions for higher headquarters commanding all units
Unit Performance Categories	• Leadership* • Command and control • Teamwork among units • Response to enemy actions • Combat outcomes • Effect of psychological stress • Effect of physical stress	• The first six items can be measured more accurately and elaborately in DIS than in field exercises for unit operations • Impact of physical environment difficult to measure in simulator-based DIS	• All items except combat stress can be derived from the appropriate units that are in each mode • Can accurately describe performance of higher headquarters commanding for all units • "Loss" of units felt in headquarters; physical attack effects are not

* Leadership is defined later in this paper. The items listed in this block have been chosen by the author as a nontechnical set of descriptive characteristics of unit performance elements that might be observed, trained, and rated in TES. The Army has a more formal structure describing its "seven operating systems," which includes Command and Control; Maneuver; Fire Support; Intelligence; Air Defense; Mobility/Counter mobility; and Combat Service Support. This more formal structure leads to the breakdown of military activity into tasks and subtasks that are rated in diverse training activities.

V. STATUS OF THE QUANTITATIVE EXPRESSION OF THE MILITARY VALUE OF UNIT TRAINING

Note: With one exception (Gorman, 1990), all the sources of data and information described in this section were formulated and had their major inputs during the period when the Soviet Union was viewed as the major military threat to the United States and its allies. Therefore, the data are, of some necessity, oriented in that direction even though, at the current writing, the "Soviet Union" has been defunct for over 4 years. While this may raise questions about the relevance and applicability of the data, it is important to remember that the data deal with combat and training for combat that are independent of scenario or of whose weapon systems may oppose U.S. troops, and they deal with the proficiency of U.S. troops and units to use their equipment in combat, regardless of the opposition. Consequently, the Cold-War orientation of sources is not viewed as an obstacle or a source of distortion in using and interpreting the data.

A. NATO DEFENSE RESEARCH GROUP (DRG) SYMPOSIUM AND RESEARCH STUDY GROUP (RSG) REPORTS

Contemporary work on the quantitative expression of the military value of unit training began with a question put by Earl Alluisi in 1980, following a heightening of Department of Defense (DoD) attention to cost effectiveness issues in training: Could it be demonstrated quantitatively that unit training improves the likelihood of winning battles? This question led to a 1985 NATO DRG symposium on the cost effectiveness of training (Orlansky, 1985) and a subsequent integrative study published in 1989 by the DRG's RSG-15 on the military value and cost effectiveness of training (Orlansky, 1989), both chaired by Jesse Orlansky of IDA.

The RSG-15 study assembled, analyzed and interpreted output from rich but previously dispersed literature that included attempts to quantify the effect of unit training on individual or small unit combat performance, explore the costs of training in association with training effectiveness, and gain insights about some combat capability improvements induced by training small units in the field and in SIMNET. The quantitative data and criteria in the different papers were not in commensurable terms, but the inferences drawn

from the assembled data yielded important insights about quantitative expression of the military value of training. From the wealth of data presented, coarse factors describing the force capability improvement or decrement derived from training could be inferred, and these factors could be applied to system acquisition decisions.

This work began the quantification process, and, subsequently, much of these quantitative data were incorporated into the IDA work described below. One of the case studies⁶ also shed light on the value of DIS training in relation to field training. In that study, eight M-1 tank platoons were rated in pre-training field performance of tasks deemed trainable on SIMNET. Then, four of these platoons were trained on SIMNET for 2 weeks and four were trained in the field. Afterwards, they were rated in post-training field performance on the same tasks. The SIMNET-trained platoons showed a higher increase in percentage of "go"-rated tasks after training than the field-trained platoons. These data are of interest for comparison with additional data on the subject of comparative training capabilities of the two methods (see subsection E below).

B . MILITARY VALUE WORK AT IDA

Also following from the Alluisi question and stimulated by Orlansky's work with NATO were two reports by the present author that sought quantification of the military value of training. The first (Deitchman, 1988) asked the question: Could training effects of hypothetically specified size make a difference in the outcomes of battles or wars? A computer simulation model of warfare in the NATO theater (TACWAR) that was available at IDA was used as the evaluation tool. Inputs to the basic scenario and weapon system database were changed arbitrarily by amounts derived from consultation with former senior military commanders (e.g., factors of 2 improvements in tank effectiveness and survivability). These changes were hypothesized to reflect greater force effectiveness because of training. The results indicated that some of these hypothetical effects of training that could be considered within the feasible realm, especially for tanks and tactical attack aviation, could make a difference in the outcome of the NATO/Warsaw Pact war that the simulation modeled. In particular, a factor of 2 improvement in effectiveness and survivability of tank and of tactical air-to-ground forces would cause a reversal of the base-case Warsaw-Pact victory in the model and turn it into a NATO victory (as measured by FEBA movement).

⁶ Included in Orlansky (1989), based on a U. S. Army letter of 25 April 1988.

The second paper (Deitchman, 1990) explored the relative cost and effectiveness outcomes of training and of hardware modifications to improve force capability for tank combat at the platoon level and for tactical bombing at the squadron level. A number of sources of relevant data generated over a 20-year period (many of them described in the RSG-15 report) were obtained and combined in the analysis. They included cost-effectiveness analyses of tank and aircraft systems; results of tank unit training exercises in the field in Europe (described in a different context by Gorman in a paper to be discussed below); results of SIMNET exercises, obtained for other reasons than training; results of a study of the effectiveness of the Service Flying Hours Programs, being performed at IDA under DoD sponsorship; and Service studies of the effectiveness of specific bombing and other live firing activities. To make the comparisons, contemporary cost analyses were combined with effectiveness data from the variety of exercises and simulations, training effects analyses, and system cost-effectiveness analyses that made up the database. Although the effectiveness data were disparate and had been generated over many years for diverse purposes, the combined analysis showed that at the unit level examined, in both the armored combat and TACAIR cases, the improved hardware or the intensified training yielded roughly comparable improvements in force capability at roughly comparable cost. Specifically:

- The 10-year cost of the M-60A1 tank upgrade to the M-60A3 and the 10-year cost of one NTC-like training session per year were approximately equal for a tank platoon.
- Either the tank upgrade or the training would lead to about a 35 percent improvement in tank platoon effectiveness in combat.
- From the available analytical (not training) results, to achieve a factor of 2 or more effectiveness in combat would require changing from the M-60 to the M-1 generation of tank, with training status in both assumed equal. The 10-year cost to re-equip a platoon with M-1 tanks was estimated to be only about 30 percent higher than the 10-year cost of the M-60A3 upgrade because operations and maintenance cost, rather than hardware acquisition cost, was the largest component of the 10-year cost total.
- The 15-year cost of increasing an A-10 or an F-16 squadron's crew flying hours, including tactical bombing practice, from 10 to 40 hours per month was roughly the same as the 15-year cost of upgrading the squadron from the A-10 to the F-16 aircraft (recognizing that the two aircraft carry out different missions).
- Either the training or the aircraft upgrade increased tactical bombing accuracy by nearly a factor of 2. However, if the training effect were used to improve

squadron performance with the more capable aircraft, the overall improvement with *both* the aircraft upgrade and the increased training could be as much as a factor of 3 1/2 to 4.

It is of additional interest to note that significant exceptions to the apparent rough equality of cost for equally effective training or equipment upgrades were retrofitting existing aircraft with advanced bombing systems and training tank platoons in a DIS environment (i.e., SIMNET). Both would entail significantly lower cost than the alternatives (buying new and improved aircraft or training with tanks at the NTC). Retrofitting an improved bombing system to an existing aircraft might cost in the neighborhood of 10 percent to 15 percent of the cost of replacing the aircraft with a new generation of aircraft.⁷ Training in a DIS environment could cost as little as 5 percent as much as training with real equipment in a training range environment like NTC.

For the armor case, the effectiveness improvements because of training or of modernizing the existing hardware were about one-third the magnitude needed to change the course of the war in the model runs of the first analysis. In the TACAIR case, the improvements were of approximately the desired magnitude. (In both cases, survivability improvements were important, but quantitative data in this area could not be found.) In the armor case, the available analytical data suggested that a new-generation tank would have produced the necessary effectiveness improvement at not much more 10-year cost than the current-tank upgrade. However, within the uncertainty bounds of the effectiveness data, the results from these comparisons suggest that hardware improvement *and* unit training in the use of this improved hardware (to ensure performance at the design level) would be necessary to ensure winning the NATO/Warsaw Pact war modeled by TACWAR.

There was a serendipitous additional result of these analyses. In both the TACAIR and the tank cases, two studies indicated that the automation of key weapon delivery equipment in new-generation systems to perform routine but demanding perceptual-motor tasks, like aim-point tracking, permitted much earlier achievement of high levels of skill that the crews using older, less automated equipment achieved only after very extensive training. If this result were borne out by further research, it would suggest that training resource allocations could be shifted from individual (skill) to unit (effectiveness) training as forces are modernized.

⁷ There is no implication that the new generation and the retrofit are necessarily equal in performance but only that both are improvements over the original.

C. TWO PAPERS BY GEN. PAUL S. GORMAN

As an outgrowth of the NATO work, of the Defense Advanced Research Projects Agency's (DARPA's) pursuit of the distributed interactive simulation that became SIMNET, and of his own considerable experience in training activities, Gen. Paul S. Gorman (USA, Ret.) prepared two papers at IDA that also bear on the "military value" question. The first (Gorman, 1990) reviews the concept of military value of training and expresses training effectiveness in terms of the expression for readiness noted earlier: $R = f(W, P, T)$, where W (weapons effectiveness) and P (individual unit proficiency) can be measured objectively, but T ("technique"), which involves tactics, weapon employment, leadership, and other intangibles, must be assessed through judgment. A significant portion of the paper describes the author's experiences in field training when he commanded the 8th Infantry Division (Mechanized) in Germany and instituted the first realistic field training exercises for armored units (REALTRAIN), which became the prototype for the NTC. His review of training exercises suggests training requirements to achieve readiness for the kind of defensive armored battle anticipated in the NATO Central Region, and retention time of the capability once achieved. Ratios between values of T, which are obtained after certain amounts and kinds of training, are adduced. T is obtained from the judgments of the training referees.

One set of results was obtained for infantry squads and for armor/anti-armor teams including (each) a tank platoon, a TOW section, and an artillery forward observer. In both cases, units trained by realistic TES, which is based on mock combat against an intelligent enemy in real or simulated terrain, improved drastically in effectiveness after the training—the infantry squads by factors of 2 and the armor/anti-armor teams by factors of up to 6. "Conventionally-trained" units (in the 1975 training idiom, training not involving TES) did poorly in comparison.

A second set of results was obtained by two-battalion brigade task forces operating in a mix of operations and simulation. The actual battle terrain was studied, surveyed, and scouted by commanders at Brigade, Battalion, and Company levels. Afterwards, they "fought" a succession of refereed battles over the terrain, with varying missions, using a detailed game board, rules of timing, and so forth, that would have applied to actual operations. This exercise tested the planning, organizing, and tactical ability of the task force commanders. The improvement from the first to the last of four battles was between 35 percent and a factor of over 3 1/2. The greatest improvement was achieved by the command groups that were the least capable initially. Gen. Gorman states that the overall

results indicate an "approximate *doubling of effectiveness*" (emphasis in original) through the entire exercise. Another interesting comment about these exercises is that there was continual learning during the four battles. Gen. Gorman believes that the learning would have continued into a fifth battle. This is relevant to estimating the amount of simulator training that might be necessary to reach the asymptotic level of "technique" in armored combat. A related judgment, made without backup data but based on extensive military experience, was that of the U.S. Army Training and Doctrine Command (TRADOC): a tank company proficient enough to defeat a Soviet regimental attack would lose 25 percent of its capability in 3 months because of crew transfers and decaying skills.

As an aside that relates to the intangible quality of leadership, Gen. Gorman describes the counterintuitive results that he obtained in a situation of rigorous observation. In one exercise, he found that some commanders who were anticipated to be effective were not effective in winning battles because they did not have a strong ability to organize and use a staff. In contrast, some commanders of apparently indifferent ability were very effective because they had competent, well-organized staffs that were able to help them.

The overall judgment in Gorman's work that realistic unit training might improve force performance by about a factor of 2 reinforces the results obtained in the "Military Value" work described in the previous subsection.

The second paper by Gen. Gorman (1992) traces the U.S. Army's development of training, from immediately before and during World War II through the institution of close combat training techniques by Gen. William DuPuy as first Commander of the U.S. Army TRADOC. The training system instituted by General Leslie McNair in 1940 to prepare the Army ground forces to fight World War II took combat training out of the classroom and into the field. It established the sequence, still valid, of individual training, unit training, combined arms training, and large-scale field exercises. It did not create realistic, repetitive field training in mock combat at the unit level but rather concentrated the field training on the proficiency of commanders from platoon up through higher echelons and on the large-scale movement of forces. In terms of infantry squad tactics and their integration with armor, artillery, and aviation, there was a training gap for close combat that persisted through World War II, Korea, and Vietnam.

Subsequent changes that were instituted by Gen. DuPuy, beginning in 1973, recognized the growing difficulty of large-scale maneuvering in the age of environmental consciousness and population pressure on ranges, and abandoned large-scale maneuvers in

favor of close-combat tactics that were deemed more important. Realistic training in mock combat for small units was instituted, with special concentration on the gap noted previously. Tactical realism, in terms of recognizing and anticipating the actions of a responsive enemy in actual terrain, became an essential part of training. The training trends that were introduced resulted in institutionalization of the concept of TES by small units (e.g., platoon to company size) as field simulations of actual battle—the forerunner of the NTC training facility using the Multiple Integrated Laser Engagement System (MILES) to measure embedded weapon engagement outcomes.

The motivation for all these changes was to reduce excessive casualties and to achieve proficiency in using economy of force. Reducing excessive casualties was necessitated by the poor performance of U.S. forces in the opening battles of many prior wars, which, as Gen. Gorman illustrates convincingly, resulted from inadequate comprehension of small-unit engagement problems against real enemies in real terrain and from consequent inadequate unit training and higher level leadership. Achieving proficiency in using economy of force became necessary as the post-Vietnam (and now, post-Cold War) Army was (and is being) drawn down in strength while its worldwide deterrence and potential combat missions have remained large in scope.

This paper ends by describing a newly emerging combat style that the U.S. armed forces are developing. It involves rapid and effective maneuver and the application of force by combined arms, suited to modern technology and the political conditions affecting the United States in its current world position and outlook. This culminates in a listing of the important factors in current training and doctrine that must be emphasized to create and maintain the kind of Army that won in Desert Storm. In summary, they are:

- Always relying on the use of combined arms
- Responsiveness to knowledge of the enemy
- Not being predictable for enemy exploitation
- All military action to have purpose, overwhelming fire and rapid movement
- All above to be inculcated by realistic training
- Training must be continual, to ensure retention of the capability for rapid response, in recognition of force personnel turnover and that conflict situations will arise without warning; and
- Training must be so effective that realistic training feels like combat and combat feels like an extension of training.

Although Gen. Gorman's second report does not shed *quantitative* light on the military value of training, it provides essential background for appreciation of this value. It also establishes the context in which military judgment and quantitative assessment must coexist and cooperate. After reading this paper (Gorman, 1992), it becomes apparent that the culmination of the training trends of the past 50 or more years has not yet been reached. In particular, the contribution of the new, networked-computation capability represented by DIS to the realistic training sequence remains to be fully articulated. Also, the contribution of tactical air support by fixed-wing aircraft to close combat remains to be integrated into the training system of the ground forces, who experience over 80 percent of the casualties and who ultimately win or lose wars.

D. ARI ANALYSES USING NTC DATA

The NTC has been in operation since 1981, and the ARI, Monterey, has supported it by analyzing data from NTC exercises. A paper presented at the symposium on the military value of training (see Section V-A) was indicative of the kind of data that would emerge from NTC. A subsequent paper, included within the RSG-15 report, and a memorandum distributed informally at a later time (Hiller, McFann, and Lefcowicz, 1991), showed correlations on the order of 0.6 to 0.7 between casualty exchange ratio scores in defensive operations at NTC and prior OPTEMPO at home bases. There was not a similar correlation for offensive operations, but it can be speculated that this might have been because offensive operations are inherently "messier"—i.e., more variable—than defense, leading to more scatter in the data and the need for a larger sample size. Later data show correlations among performance and various determining factors in both offensive and defensive operations.

The most recent ARI publication, in the making since 1987, presents a major analysis of NTC exercises and their contributions to measuring the military value of training (Keesling et al., 1992). This work analyzes the results of battalion-level training exercises from seven brigade "rotations" to the NTC. In these exercises, the combat teams were assigned missions, and they devised Mission-Essential Task Lists (METL) to follow in carrying out the missions. The METL was supposed to be drawn from the tasks listed in the Army Mission Training Plans specified in Army Manual FM 25-100; however, on average, this was done less than half the time. Preparation for the training was designed to follow the Training Management Cycle, consisting of METL development, planning, execution, and evaluation. Of the nine "Principles of Training" contained in FM 25-100 to

assist in preparing the METL, five were of special interest in both the training exercises and the analysis of the results:

- Train in Combined Arms and Service Teams
- Train as you fight
- Use unit performance-oriented training [e.g., simulated if not real terrain, but use terrain]
- Train to maintain proficiency
- Train to maintain [equipment, in working order].⁸

Among the key results and observations from the overall analysis are the following:

- The units that performed best (i.e., had the most successes and the highest casualty exchange ratios in force-on-force missions) at NTC
 - adhered closely to the Training Management Cycle
 - adhered closely to the METL
 - paid strict attention to the Principles of Training in preparing their plans and in their training.

Units that performed the worst departed the farthest from these practices.

- There was a strong correlation (on the order of 0.7) between unit OPTEMPO and performance at NTC in force-on-force offensive operations, and in defensive live-fire operations. The best brigade in the observation sample had driven twice as many miles as the worst one in the 6 months before the NTC rotations.⁹
- Longer preparation time for leaders leads to better performance in [simulated] combat.
- Ad hoc changes in plans and non-mission-relevant tasks assigned to units and their commanders degraded performance in [simulated] combat.
- Personnel instability (especially among company and platoon commanders) degraded unit capability.

The results presented by Keesling et al. invite broader conclusions about the training process and the means to achieve the greatest military value from it. Overall (and

⁸ The other four Principles of Training are as follows: train to challenge; use multi-echelon techniques; use approved doctrine; and commanders are the primary trainers.

⁹ Although this OPTEMPO result is not the same as the one cited earlier, the two results are not in conflict. In the first case, the performance comparison is based on exchange ratio. In this case, it is based on the ratings of the O/Cs. (Per Telecon with Dr. Howard McFann of ARI.)

these are the present author's observations after reading the report, not direct conclusions in the report), it appears that adhering rigorously to a set of standard procedures to prepare for combat leads to better unit performance in combat than "*ad-hoc-ery*." This does not mean that leaders and units should not be adaptable and responsive to situations as they develop. It means that systematic planning and execution of these plans in response to situations and missions, however rapid the planning process must be, lead to better results in combat than free-flowing invention. Leaders must learn to plan and develop the tasks that are needed to accomplish their missions. They must train to accomplish these tasks and manage the training, and they must maintain their units' focus on these tasks and on the units' ultimate missions.

Implicit in the report and emerging piecemeal from widely scattered comments throughout is what amounts to a definition of effective leaders. It is worth describing because, as noted in Section II-B, leadership has been one of the intangibles in force effectiveness. Although impossible to define precisely, understanding the quality of leadership is essential to articulating the military value of training. In this study of NTC outcomes, the most effective leaders showed:

- High intelligence (as determined by scores on the Armed Forces Qualification Test)
- Foresight—the ability to plan ahead and to provide their units longer mission preparation times
- Ability to "keep their eyes on the ball" and to avoid distractions
- Good communication ability
- Ability to give feedback and reinforcement to their units
- Good organizational ability [as demonstrated by staff integration; sharing information with staff; effective use of staff; operating effectively with cross-attached units in combined arms teams; integrating standard operating procedures (SOPs) with their plans and following them]¹⁰
- Decisiveness.

Presumably, these are all individual characteristics that can be determined by adaptation of the formal methods of selection and assignment that are used by the

¹⁰ Since commanders made up their own task lists much of the time, the fact that those who adhered to the lists tended to do better than those who did not may be an indirect indicator of superior planning and organizing ability of good leaders, rather than an indicator that a correlation would exist between adherence to task lists *prepared by others* and good performance in battle.

armed forces. Although the tasks, missions, and machinery are different, a reasonable presumption is that such personal qualifications for leadership would apply to individuals in the other Services as well as the Army. Some of these qualities also may define "aces," although other qualities, such as willingness to take risks and kinds of judgment, would have to be considered. These other qualities also may apply to effective leadership of units although the ARI work has not highlighted them.

E. ADDITIONAL RELEVANT DATA

Two additional papers contain data useful in the present context.

The first is a 1989 RAND report (Veit, 1989) that evaluates the effect of AH-64 crew and unit training on combat effectiveness. Results are obtained by statistical methods designed to assess judgmental data. The data suggest an asymptotic value for the subjective measure of training effectiveness. At both the company and battalion level, when the training activity is increased from once per month to three times per month, the "likelihood of effective mission" increases only half as much per month as when the training activity is increased from one time every two months to once per month. This suggests that if enough data points were available to make a smooth curve, the highest frequency of training examined would yield greatly diminished returns and that to go higher still would yield little or no return (i.e., three training sessions per month may be the most that can be used effectively). The report does not describe the exact training activity, except to describe it as attack or firing missions. It is quite possible that the numbers of training sessions leading to asymptotic values of performance in different missions or activities would vary.

The second paper, distributed by DARPA in August of 1990 (Combined Arms Test Center, 1990), shows some effects of DIS training on the performance of mechanized infantry and tank platoons in field exercises. The tests were carefully controlled and used a reasonably large sample size (nine platoons of each type), but the training and exercise time was limited: 1 day in the pre-training exercise, 2 days of orientation, 3 days of training with SIMNET, and 1 day in the post-training exercise. (For comparison, the NTC exercises described earlier allotted weeks to months for preparation, although for a far larger and more complex force). Within these limits, the results are valuable as part of a sparse set of data that apply to the transferability of DIS training to the field situation.

In each case, the platoon mission was movement to contact with an enemy platoon in daylight. The mission was divided into tasks and subtasks derived from the ARTEP,

which provided the force performance elements that were rated ("winning" or "losing" the battle were not rated). An abbreviated list of tasks includes *tactical road march; react to indirect fire; occupy an assembly area; prepare for tactical operations; move to contact [in successive operational stages]; assault enemy position; consolidate position*. For each platoon in the exercise, there were on the order of 15 tasks and 70-100 subtasks. The numbers differed somewhat for the two different types of platoons (mechanized infantry and tank). Criteria for successful training included meeting performance criteria in the field (whether the training was in the field or in SIMNET) for all tasks and for critical subtasks. Training transfer was still considered partially successful even if one or more *non-critical* subtasks could not be accomplished, as long as *all* of the critical subtasks could be performed up to standard. Failure to perform a task or a critical subtask led to an "untrained" score in the pre- and post-training exercises. Ratings were of two kinds: go/no-go by observer controllers and self-ratings by the crews.

The results of the exercise showed statistically significant improvement in the performance of tasks and subtasks (again, winning or losing the battle was not rated). The improvement was spotty, however, with improvements concentrated in some task areas (such as reacting to indirect fire) but not in others (such as assaulting the enemy position). Also, the areas in which there was training improvement were different for the two kinds of platoons. Quantitatively, the results showed:

- In all tasks, tank platoons increased from 20 percent to 40 percent trained and decreased from 70 percent to 50 percent untrained. (The differences between the sums of these numbers and 100 percent is in partial training in performance of tasks.)
- In all tasks, mechanized infantry platoons increased from 24 percent to 50 percent trained and decreased from 53 percent to 35 percent untrained. (The increment to 100 percent again is in partially trained tasks.)

The results showed that DIS provides useful training and transfer to field performance in at least a limited set of conditions for a specific set of forces. Note that the effectiveness improvement shown is about a factor of 2 in performance of the tasks being trained. Much more research would be required at NTC to show the impact of DIS training on success rate in mock combat outcomes at Battalion or Brigade level.

These tests provided additional information that is useful for interpreting the relative value of DIS and field exercises in some aspects of unit training. Ratings by participants in the tests, including crews and officers, universally found problems with the simulator graphics, which were judged to have insufficient resolution or detail for combat. It was

implied that much of the lack of trainability observed in about half of the tasks might have been caused by the poor quality of the simulator graphics. Problems were also cited with simulator fidelity—items such as size and shape of the internal space; the feel (or lack of it) of vehicle motion; and some of the human factors engineering of the actual vehicles being simulated.

VI. COSTS OF TRAINING

The most frequently, accurately, and directly measured cost of training encompasses classroom, simulator, flight, and professional development costs for individual skill training. These are the costs included in training budget listings.¹¹ OPTEMPO, flying hour, and steaming hour costs are ordinarily embedded within O&M budgets, and pay and allowance costs of personnel are included in military personnel cost categories. TES costs for unit training, whether the activities are carried out in field exercises like those at NTC, RED FLAG, or TOP GUN or in DIS installations such as SIMNET, usually include only the costs of transporting the units and their equipment (if applicable) to the training site. Facility costs and the costs of the Operational Force at facilities like the NTC or the Aggressor Force at RED FLAG and TOP GUN are embedded in other budget categories, much as base costs for aircraft flight operations are carried in categories other than the cost per flying hour of the aircraft. For the kind of unit training discussed in this paper (to achieve "level 3, Table 1" proficiency), the cost "bins" may be within base operating and support costs or may be parts of program elements under categories such as "Force-related Unit Training." They also may be embedded in system-specific programs or larger General-Purpose Forces programs such as Divisions and Tactical Fighter Wings. The method of tracking such costs varies with Service and program.

To measure the cost effectiveness of unit training, it is necessary not only to measure performance and to obtain effectiveness measures in the field or in DIS related to winning battles, but also to be able to calculate the *total* costs of training, by whatever methods are used, for comparison with or addition to total equipment costs. In the case of equipment alone, costs accounted for include initial equipment acquisition and some specified number of years of operations and logistic support that is related to expected system service life. Some of the operating costs are, in fact, training costs. Training costs in life cycle cost calculations usually include the costs of OPTEMPO, flying hours, and steaming hours. However, all of the training costs of the units, simulator networks, and

¹¹ See, for example, the DoD "Military Manpower Training Report" to Congress for any Fiscal Year.

the operation of field installations where the training takes place must be calculated and then included as part of the cost of fielding and maintaining a capability involving new systems and the forces using these systems. This is not usually done.

In some COEAs, life cycle costs are calculated from planning factors applied to the hardware acquisition costs. These factors certainly encompass OPTEMPO, flying hours, and steaming hours, but they may or may not include (accurately) the system-specific costs of unit training activities in field exercises or in DIS. Since there is a tendency to gain experience with new systems before ascertaining unit training needs, the planning factors tend to reflect prior experience rather than prospective need. This can be an important distortion if the new hardware is revolutionary enough to cause changes in tactics and doctrine that require the retraining of existing forces and then continuation of the ongoing training of future forces at different levels and with different approaches. To put the system-related and training cost algorithms on a comparable basis for all of the Service COEAs that precede system acquisition and force planning decisions, military budgeting methods will have to be rearranged so that the costs of unit training can be determined and incorporated uniformly and reliably.

VII. SIGNIFICANCE OF THE AVAILABLE "MILITARY VALUE" DATA

A. FORCE IMPROVEMENT OR CAPABILITY REDUCTION

1. Effectiveness Improvement or Decrement

The "military value" work at IDA suggested that "factors of 2" improvements in critical areas of combat (specifically, in the areas of tank combat and tactical air-to-ground effectiveness) can reverse the course of a war, that less improvement may not have the desired effects on combat capability, and that more improvement would give more assurance of success. The force capability improvement factors that can be achieved through training are not known in any exact way, but the available data, though sparse, highly scattered, and based mainly on judgment-derived scoring systems, suggest that factors of 2 or possibly better might be achieved through training. The following specific statements of effects are not unreasonable in light of the data:

- Realistic training (what Gen. Gorman has called TES—see Section V-C) at levels from platoon to battalion can improve ground force combat capability, on average, by factors of 2 above the capability that routine peacetime training and other military activity generate. The improvements range from about 30 percent for units that have high capability already to factors of as much as 3 to 6, for units that are starting very low.
- Realistic training consisting of squadron level tactical bombing practice can increase tactical bombing accuracy (and therefore effectiveness) with particular aircraft by about a factor of 2, at the cost of increasing flying hours per month by as much as two to four times. Improvements in survivability are also needed to ensure the maximum contribution of tactical air-to-ground training to winning a war. These improvements may be achieved through exercises like RED FLAG, but research is needed to confirm this.

In military combat areas essential to battlefield success, equipment replacement can also achieve factors of 2 or better improvements in force capability. Unit training effects add to these improvements. Conversely, it can be considered that:

- If there will not be adequate unit training to achieve and maintain peak unit combat skills with new equipment (i.e., the "design" levels of performance), then the performance of the equipment should be discounted by a factor of 2 in assessing a new system's cost-effectiveness. This is a rough approximation in the absence of more precise estimates applicable to equipment of particular kinds and missions.

Actually, there will probably always be some level of unit training associated with combat units. The issue here is as follows: How much is essential to achieve and maintain peak combat skills? New systems may require different amounts of unit training, with changes in associated unit training costs. Costs such as major-force retraining for new tactics and doctrine must be included. The systems being replaced may have been subject to unit training at the required level, with assignment of attending budgets, or added unit training costs could be needed for the new or for the current system if unit training with the current system was not at appropriate levels. Thus, the discounting "rule" must be applied judiciously to account for the history of the existing system and the anticipated demands of the new system. However, the rule offers more than has been available until now in accounting for the effect of training on system cost and operational effectiveness.

2. Peak Effectiveness

- From the data, asymptotic values of training in armored combat (i.e., the values where further training yields no returns) may occur after four or five exercises in a training session in the field (and, possibly, in a distributed simulator network). They may be reached at three training sessions per month in helicopter anti-armor combat. Training effectiveness is retained for a few months. According to one military judgment, capability may have declined by 25 percent or more in 3 months, suggesting a training renewal cycle time, but this is not addressed by the existing data. Maintaining a training-like activity (such as OPTEMPO) may extend such retention time, but this speculation also remains to be confirmed. Presently, it is based on the demonstrated beneficial effect of OPTEMPO on success rate at NTC. The available data on asymptotic values of training effectiveness *suggest* a realistic unit training cycle of at least once per year, but this is speculative. The data do not explicitly address the desirable frequency of the training cycle.
- From the data, asymptotic values of tactical bombing effectiveness appear to depend at least as much on career flying hours in a specialty than on immediate practice. This level is achieved after about 1500 career hours in the specialty. Bombing accuracy at this career stage is nearly twice as good as that at the beginning of the career specialty, e.g., at 300 hours. According to the data in

the supporting Air Force study (Cedel and Fuchs, 1986), the factor of 2 gain with immediate practice (flying hours per month) adds to the career level of achievement.

- Retention of peak effectiveness at the unit level is a function of skill retention and crew turbulence (transfers, rotation, training regime changes). Crew turbulence may be reduced as a result of U.S. reduction in overseas deployments. The impact on the retention of unit capability after unit training should be a subject for future exploration and tracking as the situation changes.

3. Leadership: A Key Factor in Unit Effectiveness

- It appears that the qualities of good leadership can be defined from data emerging from TES and should be predictable for individuals by methods adapted from those developed for personnel selection and assignment. The definition should be refined, specific tests should be devised, and appropriate assignment practices should be undertaken. Currently, the leadership selection process is informal. There would be value in formalizing it to increase the success rate in selecting and assigning potentially high-quality leaders from among those entering the military in times of reduced force levels.

B. UNIT TRAINING COST AND INTERACTION WITH EFFECTIVENESS IMPROVEMENT

The available analyses show that unit training costs and hardware replacement costs to achieve the desired "factors of 2" force effectiveness improvements are roughly equal (within, e.g., 30 percent or less) over system life cycles of 10 to 20 years. The magnitude of training costs arises not from the cost of individual training sessions, which even at their most expensive are considerably cheaper than hardware upgrades at force level. Rather, the high training costs result from having to sustain training to maintain unit combat capability and from having to retrain units as their personnel and leadership rotate. Thus, looking back to Figure 1, while training can increase capability quickly, the cost to sustain this new capability appears, from the analyses available, to be roughly equal to the cost of increasing force capability by replacing a force's systems with newer, more modern ones. Further, if funds for sustaining levels of unit training cost are not budgeted, then the available data suggest that force capability *with either new or existing systems* will not be maintained within roughly a factor of 2.

The fact that hardware life cycle costs and unit training costs were roughly equal in the few cases that were examined is, to the first order, a coincidence of the numbers. No established relationship between hardware and training costs is implied. However, a

relationship between unit training and life cycle costs could exist. Just as unit training cost derives from the need to sustain one kind of training, a substantial fraction of system life cycle cost derives from O&S costs, which include other kinds of training during system operating hours. It is generally accepted that life cycle cost, including O&S, will be about 1 1/2 to 2 1/2 times as high as hardware acquisition cost alone, and this is often used as a rough rule of thumb in new system planning. The results described in this paper about conducting unit training and sustaining the capability it achieves imply another rough rule of thumb: that system life cycle cost estimates must be increased by about another factor of 2 to obtain the total cost of keeping a system and its personnel in the force at peak capability.

Some kinds of hardware improvement and some kinds of training are considerably cheaper than the rough equality discussed above might indicate. First, some system improvements, such as aircraft or tank subsystem upgrades, may yield nearly or even more than the desired factors of 2 improvement in force effectiveness, at significantly lower cost than full system replacement. (This does not necessarily imply that improvements effected by the subsystem upgrades yield all the performance or operational advantages of full system renewal.)

Second, in the area of TES, DIS cost is significantly less than field training cost. However, the relative training effectiveness of the two approaches is not known. It is only hinted at in the available data. The available data (e.g., Section V-E) suggest that training effectiveness is related to simulator physical characteristics, among other things. The ratings cited earlier suggested that physical problems with the simulators in SIMNET detracted from their value as a training system. In evaluating complaints about simulator realism based on SIMNET tests, it must be remembered that SIMNET was built as an experimental and demonstration tool aimed primarily at perfecting the computer networking technology. The problems identified in association with training data should be "fixable," but costs for fixing them will vary. Given the advances in computer capacity and display technology since SIMNET was created in the mid-1980's, higher resolution and more detail in the simulators should be achievable at little or no cost beyond the development cost of the system improvements. Improving simulator physical fidelity and giving the simulators more "road feel" will add capital costs, and this may be of concern for a large assembly of simulators in a distributed training system. If fixing the problems improves the training capability of the DIS and if the probable savings in training costs suggested by the few available analyses (see Section V-B) are realized, then the investment in improved

DIS equipment to approach a particular level of capability would more than pay for itself in a fairly short time in training cost savings.

For training purposes, high-fidelity flight simulators and embedded simulation modes in fielded weapon systems should, as suggested by available analyses of simulator effectiveness (e.g., Orlansky and String, 1977), also provide training improvements in critical combat skills when they are networked and used in some sort of realistic TES. The critical question here, as with the ground forces simulator network, is whether the capabilities suggested by the comparison in Table 2 would be trained using the simulators. The question can only be answered by more research, of the kind carried out with platoons in SIMNET and at NTC, extended to the environments of all the Services and to higher-echelon forces.

The data from NTC (see Section V-D) suggest that there may be a relationship between the way the training is planned and the criteria by which the training is judged. That is, if the desired outcome is task performance, then one DIS training effectiveness result may emerge. If it is success in combat by some combat-related measures, then another result may appear. While DIS training cannot substitute for field training in some respects, in other ways it is a better medium (as suggested by Table 2). If DIS is used in combination with field training for a substantial fraction of the training needed to raise unit effectiveness levels initially and then used for much of the necessary unit skill maintenance training, there will be considerable potential savings of unit training costs.

At present, there are insufficient data to be able to relate DIS and field training quantitatively for planning and budgeting purposes. Such data are essential to planning unit training activities and controlling their costs. Therefore:

- Research to relate unit training effectiveness in DIS and in field training should be a matter of highest priority. The training cost implications and, therefore, the maintenance of force capability at high levels depend on the results of such research.
- It must be understood that OPTEMPO cannot easily be traded for unit training costs in TES. (This is distinct from the possible benefits of OPTEMPO in maintaining some level of unit capability once that has been achieved in TES.) Rather, the available data suggest that OPTEMPO (or its equivalent in flying hours or steaming hours for the other Services) is necessary as a forerunner to TES in DIS and in the field to help achieve the improvement that can be derived from realistic—TES—training.

C. IMPACT OF UNIT TRAINING ON FORCE PLANNING: BROADER ISSUES

Quantitative understanding of the cost of unit training and of the effects of unit training on combat effectiveness can help defense planners deal with broader issues than the immediate questions that might be raised about cost effectiveness of specific system decisions. This knowledge can help answer broader planning questions about improving the proficiency of current and future forces, such as the following:

- How much to spend for fast ramp-up of force capability through unit training.
- How much to spend to maintain the force combat capability achieved.
- How much to spend to keep reserves proficient or at a level where a quick surge in training can yield useful results for rapid engagement of the reserves.
- How to balance investments in equipment fixes together with unit training, compared with acquiring entire new systems, to ensure capitalizing on the improved force capability that the new systems might make available.

The results obtained from this review, sketchy as they may be, do shed some light on possible answers to these questions. Being able to answer them with even a cautiously applied quantitative assist should lead to somewhat more refined military budget allocations than those that can be reached by planners' intuitive or arbitrary judgments alone, as so many in the past have been.

The questions and brief answers that it is possible to give based on the results of this survey of the state of understanding of unit training effectiveness and cost, are listed in Table 3.

Table 3. Implications of Available "Military Value" Data for Force Planning

Issue	Answer, from Available Data
How much to spend for fast ramp-up of force capability through unit training.	The cost of appropriate training appears, at first approximation, to be very roughly equal to the cost of a hardware upgrade to achieve equivalent capability increase over the same life cycle. If a new combat system were, typically, to cost \$10 billion to \$40 billion over a 20-year period, this approximation would imply a one-time, force-wide cost of \$500 million to \$2 billion for one intensive training period in 1 year, à la Desert Storm, to upgrade a force by the same amount that the hardware upgrade might achieve. This is not a trivial amount, and <i>it does not account for training cost to sustain the capability or renew it after personnel rotation</i> . In addition, the amount will vary with the size of the force being trained. The limits given are coarse approximations, at best, but no other numbers are available.
How much to spend to maintain force combat capability achieved.	Inverting the reasoning of the previous paragraph, the full \$10 billion to \$40 billion will be required, in addition to the system life cycle cost, to maintain the capability over the time period that is analogous to a hardware system life cycle. This added cost could be lowered, at the cost of increased risk to national security, by acquiring the hardware but deferring some of the unit training. Full capability with the hardware will not be sustained, but training can be increased to achieve the capability at a relatively modest one-time cost. The price will be paid as time to achieve the capability, with reduced force readiness to meet crises.
How much to spend to keep reserves proficient or at a level where a quick surge in training can yield useful results for rapid engagement of the reserves.	Reserves are typically not trained to the level of proficiency of the average (or perhaps even the worst) of the full-time active forces. The expenditure level for (and kind of) unit training necessary to keep reserves at the proficiency level of even the lowest quartile of active forces is unknown. The amount needed to upgrade reserves to active-force unit training readiness levels will therefore certainly be greater than the amount, suggested above, that is necessary to train active forces to achieve a desired proficiency improvement. It probably varies with Service and military specialty. For example, pilots who practice individual skills in a squadron setting may be at a higher proficiency level in relation to active forces than tank crews who may have little opportunity for realistic training with the higher commands they will join in actual warfare. Thus, at best, the available data provide a lower limit to the desired costs. The lower limit is of the same order as typical, total annual manpower and training budgets for the reserves (on the order of \$1 billion to \$2 billion per year), so that the expenditures to achieve the capability levels being considered would be viewed as large.
How to balance investments in equipment fixes together with training, to ensure capitalizing on the improved force capability that entire new systems might make available.	The apparent tradeoff implied by the question really does not exist. The cost of unit training must be borne whether an existing, new, or upgraded system is emphasized in the force. The difference in unit training costs among the existing, upgraded, or totally new system may be of second order. Therefore, the decision whether to upgrade or renew begins as a function of what the hardware is expected to achieve. To achieve improved force performance above average sustained level with any of the hardware options will require added unit training costs. If new hardware demands new tactics and doctrine, unit training costs will increase even further.

VIII. INCORPORATING "MILITARY VALUE" CONSIDERATIONS IN SYSTEM AND FORCE ACQUISITION DECISIONS

Based on current knowledge, what can be done to incorporate quantitative consideration of the military value of training and of military hardware, on a comparable basis, in system and force acquisition decisions?

The current state of knowledge allows the two coarse approximations described previously:

- Adequate amounts of realistic unit training may increase force capability by a "canonical" factor of 2 from average levels, or a deficit in such training may cause a force effectiveness degradation by up to 50 percent from levels planned according to hardware design system effectiveness considerations alone
- The cost of such training over a time period about the length of a system life cycle may be about the same as the cost of incorporating a new, major weapon system in the forces.

These approximations allow rough planning at the level described in Table 3 and to be described further immediately below.

To be more precise than these simple indicators allow, a better understanding of the relationship between training input and performance output must be developed. To develop such a relationship, deeper knowledge of the training and cost components is needed. Two areas of research must be pursued:

- The validity of DIS and field training to represent the results of field operations must be established in terms of winning engagements by units of various size (up to and including, say, battalion or squadron level),¹² to allow DIS and the

¹² Gen. Gorman makes the point that the necessary large areas of unoccupied terrain for exercises involving larger units are simply not available in peacetime. Larger fleet exercises may be possible, and presumably in times of serious threat to national security, such as existed in 1941-42, the larger areas could be made available. *Instrumented* ranges might be limited in size even in those circumstances, unless the instrumentation were limited to vehicles rather than the background—e.g., use of the Global Positioning System (GPS) rather than ground stations for position location—because of their cost.

exercises to be used as tools to quantify the effectiveness aspects of unit training

- The costs of unit training in TES, for both field exercises and DIS, must be aggregated and considered as budget categories that can be managed in association with system and force budgets, if that is not done now, to be able to quantify and plan the costs of achieving and maintaining force capability at certain levels.

In the system acquisition decision area, once quantitative unit training effectiveness outputs of exercises, distributed simulations, and base operations are available for forces of particular kinds and sizes and when more accurately defined costs of unit training are available, it should be possible to enter the resulting information into COEAs for new system acquisition decisions. The comparisons made in the kinds of computer-based simulation models that are used for COEAs tend to be highly stylized. For example, model runs undertaken in planning for Desert Storm would have predicted significant casualties on the Allied as well as the Iraqi side. This is because the models are essentially resource exchange models and do not allow for variability in human performance—the effects of surprise and shock on the ability of the attacker to reduce his casualties or the ability of such attack tactics to cause the defender to break and run instead of staying to defend. However, the entry of unit training effectiveness considerations and cost estimates into such models would, at the least, put the estimates of hardware system and unit training cost effectiveness on comparable and more realistic bases than the current approaches to such estimates allow. Once the data are available, it is also possible to temper them by military judgment borne of experience (i.e., of the kind illustrated in the Gorman reports). The judgments can be used to indicate how human-related factors affect the model results and how those results should be interpreted in consequence. Such applications of military judgment to cost effectiveness analyses of military systems should lead to more credible COEAs in this time of scarce resources.

The importance of the variations in training effectiveness with time must be emphasized in the application of training cost and effectiveness data. Since COEAs seek to maximize force output measures, training effectiveness must approach the asymptotic value. Thus, the costs and effectiveness of the two phases of unit training—training to achieve peak capability and training to maintain this capability—must be accounted for. For each of the two phases, this will probably mean separate analyses of effectiveness levels and cost levels and the means to ascertain necessary expenditures. The available data suggest that TES unit training should be renewed at least once per year to sustain high unit

capability. Current practice exposes units to TES much less frequently than this. DIS, provided its efficacy were established, would be a better tool than field exercises (or OPTEMPO and its analogs in the other Services) to do this on a systematic basis simply because it is far less expensive than the other methods. The results would have to be integrated with the outputs of some (currently unknown) number of field exercises, both to capture the effects of factors not included in DIS and for credibility with the users of COEAs.

In the force planning area, once unit training costs can be calculated, it would become possible to allocate and manage a total budget that includes provision for unit size and training readiness. If more powerful forces are needed, apart from the obvious possibility of force expansion, which must entail increased budgets, extensive time, and a degree of political turmoil, changing allocations within a given overall budget can allow extensive force planning versatility. Training (as well as equipment) dollars could be traded for dollars used to sustain forces of different capabilities. For example, if the costs were known, it would be possible to allocate unit training funds to maintain one of the following:

- A relatively small but highly ready force able to go into combat on short notice
- A large force requiring additional (e.g., "topping off") training to go into combat
- A force with some units, in numbers allocated by plan, in each status.

Planning for forces to be able to undertake combat operations on short notice is essential in the new strategic environment for national security, and it cannot be accomplished without taking training costs into account. Giving visibility to the unit training budget would give an element of insurance against one of the tendencies to allow a force to go "hollow" through the budgeting and funding allocation process.

In the absence of the necessary further research results, the rough "rules of thumb" permitted by current knowledge can give a coarse approximation of system and force cost and effectiveness results. Even such coarse estimates should be improvements on the way the cost-effectiveness estimates are made today, without accounting for unit training cost and effects.

REFERENCES

Alluisi, E.A., "The Development of Technology for Collective Training: SIMNET, A Case History," *Human Factors*, Vol. 33, No. 3, June 1991, pp. 343-362.

Cedel, LTC T.E., and Fuchs, LTC R.P., "Factors Affecting Pilot Proficiency," Study Report, Fighter Division, Air Force Center for Studies and Analyses, December 1986.

Combined Arms Test Center, Headquarters TEXCOM, "Close Combat Tactical Trainer (CCTT), Force Development Testing and Experimentation (FDTE), RCS ATTE-3," Final Report, TCATC Test Report FD 0200, Fort Hood, Texas, August 1990.

Dayan, M., *Diary of the Sinai Campaign*, Schocken Books, New York, 1967.

Deitchman, S.J., "Further Explorations in Estimating the Military Value of Training," IDA Paper P-2317, January 1990.

Deitchman, S.J., "Preliminary Exploration of the Use of a Warfare Simulation Model to Examine the Military Value of Training," IDA Paper P-2094, March 1988.

Gorman, Gen. P.F., (USA, Ret.), "The Secret of Future Victories," IDA Paper P-2653, February 1992.

Gorman, Gen. P.F., (USA, Ret.), "The Military Value of Training," IDA Paper P-2515, December 1990.

Hammon, C., and Horowitz, S., "Relating Flying Hours to Aircrew Performance: Evidence for Attack and Transport Missions," IDA Paper P-2609, June 1992.

Hammon, C.P., and Horowitz, S.A., "Relating Personnel and Training Resources to Unit Performance: Identifying Data on Performance in the Military," IDA Paper P-2023, September 1987.

Hiller, J., McFann, H.H., and Lefcowicz, BG L.G., "Does OPTEMPO Increase Unit Readiness? An Objective Answer," ARI Field Unit, Monterey, April 8, 1991.

Horowitz, S.A., Hammon, C.P., and Palmer, P.R., "Relating Flying Hours to the Performance of Aircrews," IDA Paper P-2085, December 1987.

Keesling, J.W., Ford, P., O'Mara, F., McFann, H.H., and Holz, R., "The Determinants of Effective Performance of Combat Units at the National Training Center," Final Report, Vol. I, Findings, published by U.S. Army Research Institute, June 1992.

McFann, H.H., "Relationship of Resource Utilization and Expenditures to Unit Performance," Presented at NATO Defense Research Group (Panel 7) RSG-15 Workshop on the Military Value and Cost Effectiveness of Training, Cambridgeshire, England, 10-12 October 1990.

Orlansky, J., and String, J., "Cost Effectiveness of Flight Simulators for Military Training," IDA Paper P-1275, August 1977.

Orlansky, J., ed., "Proceedings of the Symposium on the Military Value and Cost-Effectiveness of Training," NATO Defence Research Group, Panel 7, on the Defence Applications of Operational Research, Brussels, 7-9 January 1985.

Orlansky, J., ed., "The Military Value and Cost Effectiveness of Training," NATO Defence Research Group, Panel 7, on the Defence Applications of Operational Research, RSG-15 on the Military Value and Cost Effectiveness of Training, AC/243 (Panel 7/RSG-15) D/4, Brussels, 28 December 1989.

United States Department of Defense, "Conduct of the Persian Gulf Conflict: an Interim Report to Congress," July 1991, pp. 21-2, 3.

United States Military Academy, Department of Social Science: "Are Smart Tankers Better Tankers: AFQT and Military Productivity," December 1984.

Veit, C.T., "Effects of APACHE Helicopter Crew and Unit Training on Combat Mission Effectiveness," RAND Report P-7590, August 1989.

Wainstein, L., "The Relationship of Battle Damage to Unit Combat Performance," IDA Paper P-1903, April 1986.