Old Dominion University

ODU Digital Commons

Computational Modeling & Simulation Engineering Theses & Dissertations

Computational Modeling & Simulation Engineering

Spring 2008

Developing a Repeatable and Reliable Methodology to Determine Return-on-Investment

Mark Cullen Nesselrode
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/msve_etds

Part of the Military and Veterans Studies Commons, Operational Research Commons, and the Systems Engineering Commons

Recommended Citation
Nesselrode, Mark C.. "Developing a Repeatable and Reliable Methodology to Determine Return-on-Investment" (2008). Doctor of Philosophy (PhD), Dissertation, Computational Modeling & Simulation Engineering, Old Dominion University, DOI: 10.25777/d5x8-g178
https://digitalcommons.odu.edu/msve_etds/20

This Dissertation is brought to you for free and open access by the Computational Modeling & Simulation Engineering at ODU Digital Commons. It has been accepted for inclusion in Computational Modeling & Simulation Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

DEVELOPING A REPEATABLE AND RELIABLE METHODOLOGY TO DETERMINE RETURN-ON-INVESTMENT

by

Mark Cullen Nesselrode
B.S. Mathematics June 1976, United States Naval Academy
M.S. Mechanical Engineering March 1985, U.S. Naval Postgraduate School
Degree of Mechanical Engineer March 1985, U.S. Naval Postgraduate School

A Dissertation Proposal Submitted to the Dissertation Committee
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ENGINEERING WITH A CONCENTRATION IN
MODELING AND SIMULATION

OLD DOMINION UNIVERSITY
May 2008

Approved by:

John A. Sokolowski (Director)

Michael L. McGinnis (Member)

Charles B. Keating (Member)

Lawrence C. Schuette (Member)

ABSTRACT

DEVELOPING A REPEATABLE AND RELIABLE
METHODOLOGY TO DETERMINE
RETURN-ON-INVESTMENT

Mark Cullen Nesselrode
Old Dominion University, 2007
Director: Dr. John A. Sokolowski

The high costs of modern weapons systems, fuel, personnel, and increasing environmental awareness is forcing U.S. forces to rely on simulation for training. Further, the need to train in a robust and theater specific environment is critical to ensure operational readiness immediately upon reaching the combat theater. Finally, the rapidly changing nature of the conflicts, crises, or contingencies in which U.S. forces are involved requires a new approach to training.

The purpose of this research is to develop and demonstrate the feasibility of a methodology that can provide a reliable, consistent method for determining ROI in various forms, focusing on the staffs of Operational level commanders (those staffs who form Joint Task Force staffs or Service Component Commander staffs) so that the most appropriate format can be utilized in determining warfighting improvement or training vs. exercise cost. The major components of the ROI formulation are costs and performance assessment which can be applied in various combinations to quantify the performance benefits achieved for the expenditure of resources that are linked to specifically assessed performance parameters. The performance and cost parameters should apply to any operational or training environment, but the ability to evaluate the potential advantages of training in a completely synthetic environment as opposed to either a live or mixture of live, virtual, and constructive environments is of special interest. The cost components required to determine ROI must provide a reliable methodology to identify the costs of live and synthetic training regardless of the Service or Headquarters staff involved. A

second necessary factor is the means to assess the performance of a Component Commander or Joint Task Force staff performance during a training exercise.

The development of feasible methodologies, permitting evaluation of ROI in various formats, will provide a quantifiable means to assist in decisions regarding deployment readiness or support any Service or the Department of Defense during budgetary discussions. The ability to quantify ROI in the most appropriate or most applicable format can lead to better allocation of resources and to the ability to employ the most prepared or appropriately trained staff in a given crisis scenario.

ACKNOWLEDGMENTS

There are several groups of people who have made the completion of this dissertation possible and without whose support and patience the desire to persevere would have been far more difficult.

The staff at VMASC was always ready to assist and to offer a kind word. The trio of "Administrative Goddesses," Sheila Flanagan, Tracey Vann, and Karen Thompson lent support from travel, to room setup, to the small things such as name plates that made every interaction enjoyable and moved the process forward.

The staff at the Joint Warfighting Center, especially Tim Baker, Colonel Mike Armstrong and John Vinett supported this effort by affording access to data, offering encouragement, and providing a forum for discussion to ensure that the project met both the needs of Joint Forces Command and Old Dominion University.

There were several individuals from various staffs and commands throughout the Tidewater area that were critical in every phase of this project. The backing of Vice Admiral Chanik, Captains Steve Kelly, Steve Zaricore, Steve Snyder and Kevin Knight, Commander Mark Peot, and Lieutenant Commander Lyons were instrumental in the successful completion of the major event of this dissertation. Special thanks are extended to Vice Admiral Chanik whose friendship, vision, and faith never wavered.

Russell Williams from US Fleet Forces Command and Captain William Kovach are not only former shipmates, but their belief in the project, their willingness to either supply data, validate methods, and finally to proofread hundreds of pages was always amazing and welcome.

Captain John Kersh and Keith Barber were surprise contributors to this effort, but they also willingly and unflaggingly contributed hours of their time to assist in data collection, methodology validation, and proof reading. Their enthusiasm for this project

and the potential it held for not only the Armed Forces but Federal, State, and Municipal agencies provided insight and excitement for the project.

One individual who has not only been a friend, but advisor, advocate, proofreader, data collector, and formatting assistant is Commander Al "Hans" Kohnle. It is a fact that without his constant encouragement, support, assistance, and inexhaustible energy, this project could not have been completed. He embodies all that is meant by the words shipmate and colleague.

I have been blessed by a supportive and cooperative dissertation committee. Dr. Mike McGinnis and Dr. Charles Keating have been enthusiastic and very steady contributors and supporters of this project from its inception.

Dr. Larry Schuette is one of those rare individuals who has vision, enthusiasm, and is truly committed to the betterment of the U.S. Navy. His friendship since our first endeavor on board the USS ANZIO (CG-68) has been of inestimable value. He was the first to suggest that the PhD be undertaken, and he has stood by me in every phase, including being a member of the dissertation committee. There are very few leaders and academics of his caliber; it is a privilege to have his counsel and friendship.

Dr. John Sokolowski was the first to blaze the Modeling and Simulation PhD path at Old Dominion University. His experience and background as a nuclear trained Submarine Officer have enabled him to be phenomenally capable as both the Director of Research for VMASC and as a professor in the classroom. John has been involved in the pursuit of this degree from the very first interview and he was the one responsible for devising the Research Intern program which made completing the research phase of the dissertation possible. He has filled several roles to assist in the completion of this project: professor, mentor, friend, editor, and dissertation committee director. I am indebted to him for far more than completing the dissertation, it has been an honor to be associated with him for the last three years and I hope this association continues for many years to come.

Even with all of the support from all the individuals named above, there are those who must endure the doubt, frustration, and angst of a project of this magnitude. The Quail Point Cove support team of Harry and Kathy Acker and Bill and Beth McGloon has kept me on an even keel when I lost focus. A dinner, ball game, barbecue, golf game or just a beer have always been available when things looked a bit dark. Thanks very much.

The last and most profound thanks go out to the one group that has enabled, cheered, cajoled, humbled, and always believed in the seemingly never ending list of challenges I have embarked on. Even with their now very busy, complex, and sometimes extremely challenging lives, Lauri, Christy, and Lynne, the most wonderful daughters any person could have, have always offered encouragement and badly needed inspiration. I only hope that I have offered a fraction of the love and assistance that they have given over the last three years.

Finally, there is one person, who has waited over 20 years for this journey to end. During the long deployments, endless underway periods, and more recently evenings and weekends when, even if I was home, it was just like being at sea, my wife, Martha, has been incredible. Whether she needed to be head cheerleader, confidant, advisor, critic, motivator, or injector of some common sense, she has endured all of this effort, which came after 30 arduous years. It is for her that this work has been undertaken. It is 22 years later than she initially had hoped that the title Dr. could be achieved, but none-the-less, after all the gifts she has given me, I hope this very small one at least begins to repay what can only be seen as a life of service to all of our family, but especially to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

1 INTRODUCTION

This chapter provides an introduction and the motivational basis for developing and demonstrating the feasibility of a process to measure Return-on-Investment (ROI) regardless of training environment (which could translate to any combination of Live, Virtual, or Constructive (LVC)) for large-scale distributed training at the Operational Level of War (1)[1], (2)[2]. The high costs of modern weapons systems, fuel, personnel, and increasing environmental awareness is forcing U.S. forces to rely on simulation for training.

Further, the need to train on the systems, which everyone from a junior operator to a theater commander will use, is becoming necessary to ensure operational readiness immediately upon reaching the combat theater. Additionally, the rapidly changing nature of the conflicts, crises, or contingencies in which U.S. forces are involved requires a new approach to training. It is no longer sufficient or even efficient to train individuals as was done in World War II or even Vietnam and assume that these individuals will function effectively upon arrival (3). The widely varying nature of missions required of U.S. forces also requires innovative approaches to preparation for both the operators in the field and the staff commanding those operations (4). The rise and ever expanding use of distributed synthetic training permits training from an individual operator to the

---

[1] Citation and reference list format for this manuscript are taken from the *Journal of Defense Modeling and Simulation*

[2] The Operational level of war involves the planning, conduct and sustainment of campaigns or major operations within either a theater or specified area of operations. The activities imply a broader dimension of time and space than do tactics; they ensure the logistic and administrative support of the tactical forces and provide the means by which tactical successes are exploited to achieve strategic objectives. The focus at this level is on the design and conduct of operations using operational art .. to design strategies, campaigns, and major operations and organized and employ military forces.

Commander of a Joint Task Force (CJTF), major Component Commander (FCC)[3], and any or all echelons and staffs in between. This sort of approach is a critical element in preparing U.S. Forces for the dynamic and very broad scope of operations that are most likely or expected.

The primary purpose of this study is to quantify, in terms of costs and performance assessment, the Return-on-Investment (ROI) and specifically the benefits to be gained by conducting distributed synthetic training at the Operational or Strategic level versus that same cost utilizing live forces. As a means to quantify ROI, it is necessary to develop a consistent methodology for identifying the costs of live and synthetic training as well as identifying a means to arrive at these costs in a reliable and acceptable manner regardless of the Service or Headquarters staff involved. The other critical process in quantifying ROI is the development of algorithms that accurately assess the performance of a CJTF or FCC staff during a training exercise.

The means to develop ROI and the subordinate criteria must apply to whatever training environment within the LVC spectrum is utilized, but ultimately, the most useful ROI would permit contrasting either a live or live and virtual environment with a completely distributed synthetic environment. The specific simulation or stimulation systems used are not important, only the costs involved in utilizing these systems to conduct training are important. Suantak provides an excellent description of various systems used to train U.S. Army personnel from an individual Soldier to a Corps Commander in her 2004 dissertation. The author emphasizes at the conclusion of her examination the fact that:

> "The vast majority of military simulations focus on force-on-force combat models and simulations. These models are primarily concerned with

---

[3] For purposes of this proposal, the Operational Level staffs will indicate the functions and staff of either a Joint Task Force Commander (CJTF) or Functional Component Commander (FCC) typically comprised of a Senior Flag or General Officer from any one of the Services.

> collecting and disseminating information about terrain, weather and troop movements, engagements, and logistics. These simulations place great emphasis on low-resolutions data, modeling down to individual bullets. While providing detailed information to commanders and soldiers, this level of resolution creates an enormous amount of data that must be managed and communicated, demanding significant computer and time resources (5)."

The key here is that there is no emphasis on utilizing the combat systems that are being used in combat and on a daily basis that operators "train as they fight." Further, the ability to assess the performance of the operators at lower levels, but notably those at the CJTF or FCC level is not accounted for. Therefore, the fundamental questions remain, what is the ROI for investing in all of these systems that are increasingly sophisticated? Even though it appears intuitively obvious that simulation is a more cost effective method of conducting many types of training, especially involving large numbers of friendly and opposition forces, how much does simulation really cost? How much money is saved? Is there any compelling need for live training above the individual or small unit level? How well are the Commanders prepared for the next task required? What sorts of missions will be required that are not necessarily "warfighting" and can preparation for those be effectively achieved using synthetic training methods?

The discussion and development to follow will provide a background of some of the key work accomplished to date to attempt to quantify training performance, methods that have been proposed to formalize costs, and classical as well as contemporary views of ROI. Since this research deals with evaluation of performance, costs, and ROI in a military context, the extensions that must be considered from an industrial or corporate setting will also be examined. Significant work on each of the three areas mentioned, performance evaluation, costing, and ROI, has been completed, but the applicability of any of this work to the military, especially at the Operational Level has not been either well executed or formally implemented in a fashion that would permit senior decision makers within the Department of Defense (DoD) the ability to assess the utility of major investments to the Joint National Training Capability (JNTC) or the readiness of any CJTF or FCC staff. The necessary steps to frame this research and ultimately a useful

methodology and tool, include relevant reviews of work to date, a brief overview of the current state of the art for performance evaluation, a review of costing methodologies in use, defining the changing nature of ROI, and finally a methodology to provide a quantitatively based tool that can be used by the Senior Staffs in question.

1.1 Joint doctrine, training, assessment, and the CJTF/FCC

Prior to the mandates set forth by the Goldwater/Nichols Act of 1986 (6) U.S. forces conducted Joint and even Combined operations as separate Services or Allies with direction provided by command staffs composed of members of the involved Services or Allies, typically led by a senior officer appointed (or agreed to) by the leaders of the Joint or Combined force. Operations in both the European and Pacific theaters were conducted successfully in this manner, but there was no formal requirement to plan, train, or operate as a Joint force. The Goldwater/Nichols Act formalized the requirements for all of the above. The adoption of the Universal Joint Task List (UJTL) (1) and subsequently the Joint Training Manual for the Armed forces of the United States (7) only set forth the tasks required of all U.S forces, but methods to devise measures and standards of performance for each of those tasks.

The construct for training, especially at the Operational level of war, which is the focus of this research, can be found, at least at an initial level within the UJTL as depicted in Figure 1.

Figure 1: Joint Training and Exercise Categories (1)

The two primary categories of concern from the figure are categories 3 and 6, which are described as:

- U.S. Joint Training which is based upon joint doctrine to prepare joint forces and/or joint staffs to respond to strategic and operational requirements deemed necessary by combatant commanders to execute their assigned missions.
- Training to prepare the joint commanders and staffs, at the strategic and operational level of war, to function in a joint interagency and/or intergovernmental environment (1).

In addition to training in a Joint, Interagency, or Allied environment to achieve specific objectives or missions, there is also a desire to train to specific Joint Capabilities which might better define the requirements of the Combatant Commander (COCOM) as shown in Figure 2.

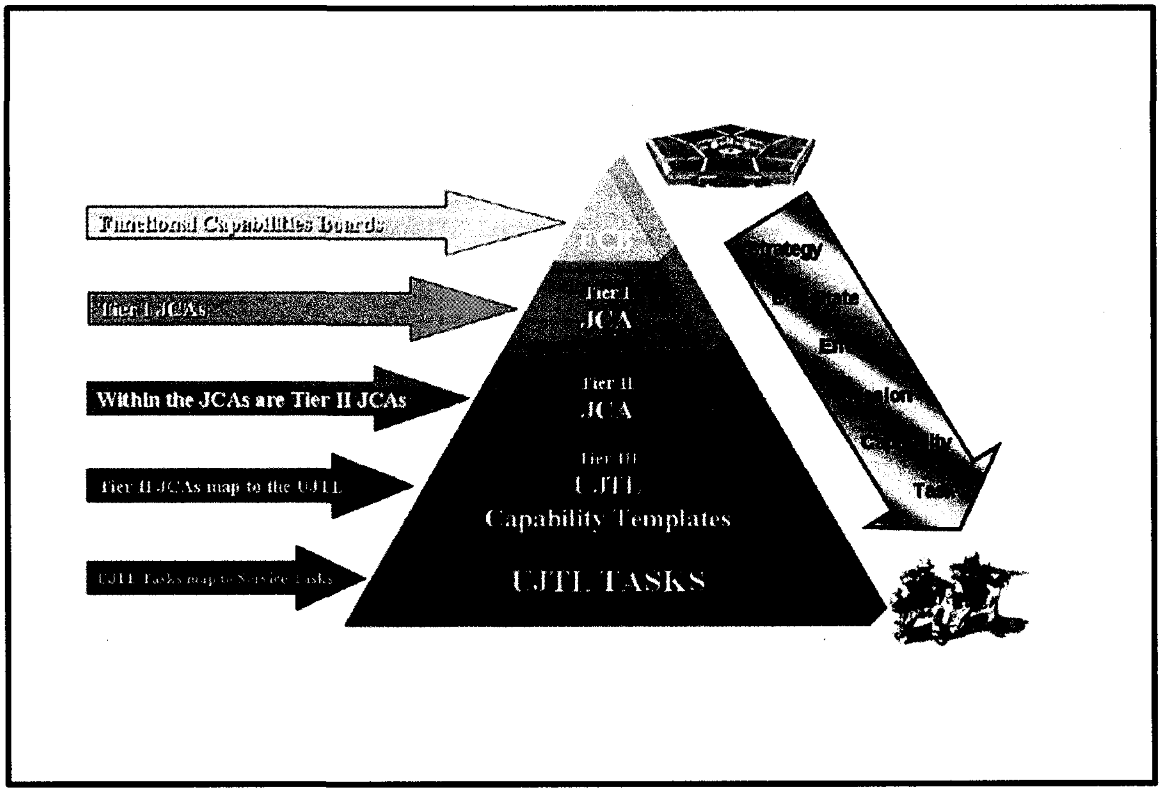

Figure 2: UJTL Relationship to Joint Capabilities (1)

Regardless of the requirement levied by the COCOM to either a CJTF or FCC, the need to prepare for the requirement in the most efficient and effective manner, and the ability to report the readiness for the required mission or capability is a pre-requisite to the inception of operations. The common language, developed as a result of Goldwater/Nichols is that of the UJTL. At the operational level, the specific operational requirements are specified in Annex C, Appendix C Enclosure B wherein the specific Operational tasks, measures and criteria are provided (1). There is recognition in various

sections of the UJTL that all tasks may not be able to be objectively observed and that in some instances the only measures that may be applied may be binary[4] in nature, but the guidance provided in Appendix B, Enclosure B states:

- Measures should focus on the outputs, results of performance, or on the process to achieve the task. In identifying dimensions of task performance, focus on the outputs or results of performance and, in selected cases, the process followed (e.g. number or percentage of sub-steps performed correctly or in the correct sequence) The dimensions of task performance should not be peculiar to a specific means for performing a task, rather, they should apply to all means that can be employed to perform a task (1).

From the preceding discussion, there is ample doctrine and guidance to build a framework available to all US forces, at any level, but specifically at the operational level, to determine what standards are necessary to train for a specified mission, to construct training scenarios, and to report the results upon completion, yet this is still not the case within the Department of Defense (DoD). The use of specific measures and the UJTL to report readiness and to specify training requirements is growing, but is still notably behind the applications used in the Services at the individual or tactical levels.

Unlike a corporate setting, there is no decision to be made regarding either the need for training or if (this is an assumption that will be more fully discussed later) the training has resulted in the requisite readiness for the anticipated operations. The issues are, what is the optimal form of that training, how much training is required to achieve operational readiness, and how long will that training take? A thorough understanding and application of both performance assessment and costs associated with the training, regardless of venue readily aid in making decisions about how to train, how long to train, and the cost of that training. More specifically, at the Operational Level of War, which is

---

[4] Throughout this proposal, binary observations will include Yes/No, or very critical measures that must be accomplished at only a given level, e.g. 100% to be considered acceptable.

one of the limiting factors for this project, is there actually a need for any training other than that provided in a synthetic venue?

Once the effectiveness of the training conducted is known, the logical question requiring attention is what did the training cost? Various methods of cost accounting exist throughout the Services, at the COCOMs, and even at some Operational Level headquarters, but how does each of the above equate the value of the training to specific resource allocations? If costs are cut, can a definable change in the level of readiness of a specific Operational Level staff be predicted? If resources are augmented, is there a methodology in place to predict what additional capabilities or operational effectiveness might be achieved? As noted in the introduction, either of these costing scenarios is possible, very certainly the former, and yet, other than viewing the impacts resulting from imposing the actions discussed, there are very few concrete answers to any of the questions posed.

The last question, which is a natural antecedent to the first two is, what was the gain (more frequently heard from very senior DoD officials or Flag/General Officers, what do we get?) from the investment or increased resources for training? Is there a tangible operational benefit to the very expensive infrastructure being used to link Operational level staffs to a live, virtual, or constructive environment? Is there any feasible way to demonstrate the additional capability that some new technology was procured to provide? In other words, what was the ROI for any of the above? As will be discussed and proposed in later sections of this document, perhaps ROI requires some re-defining, but at a minimum there must be a method to quantifiably answer the questions to senior leadership that can be used to justify the expenditure of precious funds to various leaders and bodies, including the Congress who ultimately regulates DoD funding.

The subsequent sections of this proposal will discuss what has been investigated, inside and outside of DoD to assess performance, as well as what has been investigated and implemented with regard to costing and finally Return-on-Investment (ROI). These discussions will then be followed by specific proposals that will extend the body of

knowledge in all of these three areas ultimately permitting the calculation of ROI in some specific cases.

At the outset of these discussions, it is recognized that the volume of work accomplished to date in all three of these areas is staggering, and in some cases has been revised and updated several times. It is not the intent of this proposal to conduct an exhaustive survey of all these bodies of work, but to highlight the work that specifically applies to the methodologies relevant to the development of a repeatable and reliable formulation of ROI.

2 BACKGROUND

This chapter will discuss the areas of Performance Assessment, Costing, and Return-on-Investment (ROI) as they apply to training an operational level staff. Each area will be treated separately so that previous work, the state-of-the-art, and finally the proposed contribution to the current body of knowledge are clearly explained. The discussions will necessarily focus on a military context, but a later section of this proposal will also address the potential applications outside of DoD for a similarly devised methodology.

2.1 **Performance Assessment**

2.1.1 Previous Work on Performance Assessment

There have been efforts to quantify performance assessment for many years, but a very significant investigation into performance with the possibility of utilizing simulation for large teams was conducted by McGinnis and Stone. They presented a paper detailing their research at the 1996 Winter Simulation Conference, Measuring the Effectiveness of Simulation-Based Training. The opening statement is compelling and still accurate," The relationship between measures of training performance and the achievement of training goals is often difficult to quantify (8)." The focus of the article was on teams or units versus individuals, who would be expected to use distributed or large scale simulations. The article then presents, in a logical sequence, a brief overview of previous work in training assessment, especially noting Sloman's Instructional Systems Design (ISD) model. Don Clark very succinctly defines ISD at his Big Dog website. Instructional Systems Design has several key elements:

- It is competency based in that learners are required to master a Knowledge, Skill or Attitude (KSA). The training focuses on the job by having the learners achieve the criteria or standards necessary for proper task performance
- It is sequential such that lessons are logically and sequentially integrated

- It is tracked in that a tracking system is established that allows changes and updates to the training materials to be performed efficiently
- It permits evaluation and corrective actions, which allow for continuous improvement and maintenance of training information that reflects current status and conditions (9)

Sloman's model reflects all of these characteristics and sets up a five –step process beginning with an analysis of training needs and concluding with the implementation of training (or delivery). This sequential methodology is shown in Figure 3 (10).

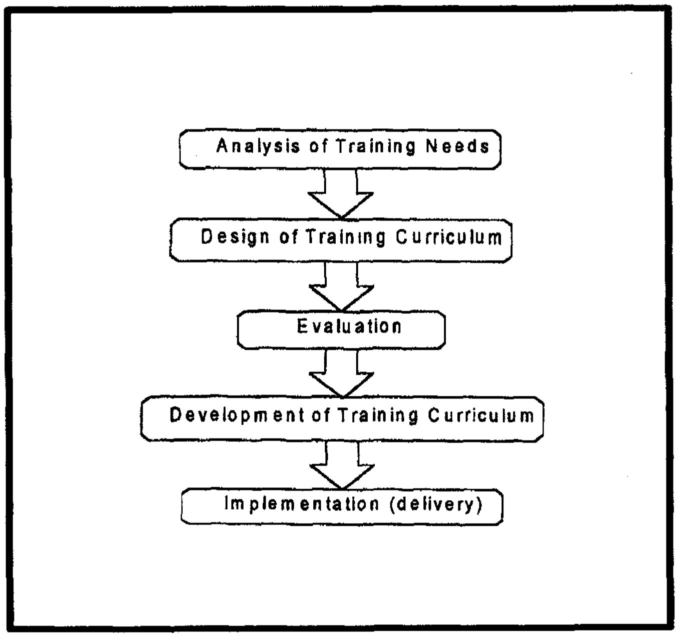

Figure 3: Sloman's ISD systematic training model

McGinnis and Stone's proposal was based on the premise that some form of assessment methodology was needed to evaluate teams or units undergoing training vice training undergone by individuals, since this level of training would be an appropriate candidate for a distributed or large-scale simulation platform. A methodology that adapted the process advocated by Sloman was presented which had six major components, including feedback. A discussion of possible evaluation standards and how

these should be used to remove bias and subjectivity led to the formulation of an elementary simulation-based training effectiveness measure.

The methodology proposed by McGinnis and Stone assumes that there would be some hierarchical scheme based upon specified tasks and that these tasks could have multiple levels of subtasks (8). They also explained that there must be an implicit understanding that any evaluation system required the existence of precise measurement standards for each task or subtask. Though the standards that are available for this proposal have not been used extensively, there is sufficient coverage of the key JMETs to permit application to any mathematical method discussed. The highest-level tasks were depicted as those requiring completion for mission success. For the purposes of this discussion, those tasks could be considered UJTL JMETs or some appropriate grouping of JMETs. The subtasks are the skills or "how" the task can be accomplished. Each subtask must be assigned a weight according to its importance to either the parent task or mission. The sum of all the weights for subtasks branching from any specific parent must not exceed one. Figures 4 and 5 provide a pictorial depiction of the McGinnis and Stone scheme (8).

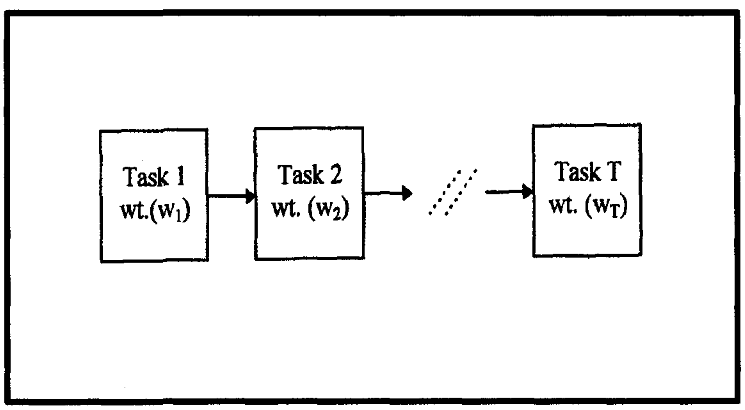

Figure 4: Upper Level Training Task Weights

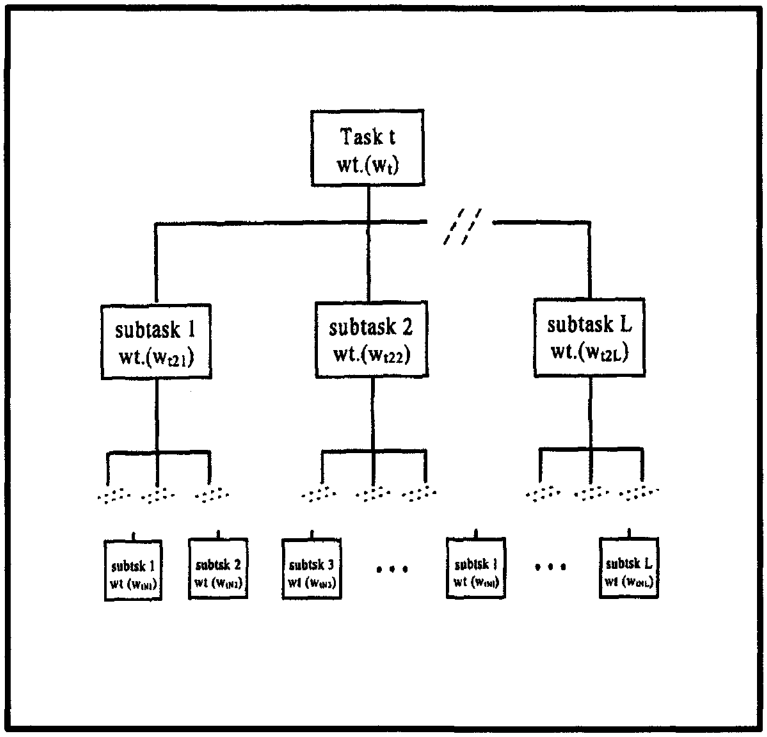

Figure 5: Subtask Hierarchy and Weights

The final "grade" or score for any mission is calculated by combining the weights (w) and scores (s) for each branch. Each branch is then combined with its appropriate weight with other tasks L, branches M (within a level), and levels N, until a final value is determined for the mission. The mathematical representation is given by the following (8):

$$\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} w_{ijk}\, s_{ijk} \tag{2.1}$$

Though only a simple assessment example was depicted, the concept shown was hypothesized to be capable of improving simulation systems to be developed, as well as the methods used for measuring training, which is a hypothesis that will be tested during the research phase of this proposal. Additionally, the application of this method to non-military applications such as training emergency personnel was addressed. The use in an Emergency Response scenario is but one possible application for the various performance assessment methodologies that will be developed during the research phase of this project; others will be more fully explored in a later section of this proposal.

McGinnis and Stone's paper was referenced in a follow-up paper that proposed training at the battalion level and higher in a completely synthetic environment might be possible when the Warfighter's Simulation (WARSIM) application was fully developed. A critical question facing the military and many industries and governmental organizations is whether fully synthetic environments can fulfill their training needs (11). The McNett et al. paper noted that use of synthetic environments could reduce the exclusive reliance on live exercises, which is attractive as environmental impacts, fuel costs, and scheduling issues are making the opportunity to utilize a L,V,C combination more attractive. In any case, the methodology required to evaluate the performance of the training audience must be established and effective.

The application of Instructional Systems Design proposed by Sloman has been refined since the McGinnis and Stone paper was presented to specifically address the issues of electronic learning (e-learning) and electronic training (e-training), but the concepts remain valid. In the updated treatment, Sloman has modified the process depicted in Figure 3 to include the necessary interactive exchange between the original five elements. This interactive approach reinforces the concluding remarks made by McGinnis and Stone regarding improving both simulation and training, since the feedback from the evaluation can be used in both development and analysis.

Kirkpatrick also uses many of the concepts that he originally proposed in the 1960s, but he has refined and expanded the discussion of assessment, especially at the team level. He has also provided insights into formulations for ROI (12).

Slightly more than ten years have elapsed since the McGinnis and Stone paper, yet there are still widely varying views regarding the progress simulation, training assessment, and training development has made. One view, expressed in an article in National Defense Industrial News by Jean touts the strides made in using simulation and assessment as a readiness tool for U.S. forces, though her discussion revolves around a Navy application (13). In an almost opposed view is a report published in May 2005 by the Army Research Institute that encourages the expanded use of simulation to more effectively train at the battalion and higher levels, which seems to echo the opinions expressed by McNett et al. and recommends more effort be put forth in developing assessment tools to make this training more effective (14). The specifics of the Holden research will be discussed in a subsequent section of this proposal.

Since the work pioneered by McGinnis and Stone, the need to assess performance, whether to gage the value of training, to estimate ROI, or to assist in developing effective training scenarios, whether these are based solely on a synthetic or LVC combination, has become increasingly important. Sloman, Kirkpatrick, Cokins, Phillips, and Kaplan have commented that the use of costing alone to determine value of any activity, but especially training on among many others is misleading. Each of the named authors has revised previous works, ((10), (12), (15), (16), (17)) in some cases a second or even third time, to account for the fact that performance assessment is a necessary component in determining the value of any activity, but especially in terms or training and ROI.

In a previous paragraph of this dissertation, the fact that the need to assess team performance is not unique to the military was discussed. Large utilities, especially commercial nuclear power plants, state and municipal emergency response centers, and even large corporations have a need to train supervisory teams, who would be the functional equivalent of the FCC or CJTF staffs to respond to a myriad of crisis scenarios. It should also be noted that although the terminology _METL is not commonly used in a non-military context, the ability to extend the use of this terminology to the aforementioned examples, if not beyond, might actually be worth considering, especially

as training between the U.S. military and other agencies expands to include Homeland Defense and Defense in Support of local Agencies. For purposes of illustration, owing to the fact that the complexity of the training issues facing the U.S. military require very complex scenarios across a broad range of capabilities, a military context will be used. Senior leaders, (such as retired General/Flag Officers) trainers, instructors, and subject matter experts would determine the appropriateness of the conditions and standards chosen. One retired General Officer who routinely assists in the training and certification of operational level staffs is General Gary Luck who works for Joint Forces Command in the J-7 (Training and Readiness) directorate as a Senior Mentor. His experience, as well as the team of Observer Trainers that support him bears out key points from several authors regarding performance assessment. As noted by McGinnis and Stone, as well as Tillson, Phillips and Cokins, General Luck's views on assessment and the value it plays in training strike a balance between the qualitative and the quantitative. Further, he believes that the Commander is primarily responsible for training and assessing his staff, and thus the Observer Trainer's main function is to assist in improving performance in areas requested by the Commander as well as to validate the performance of the staff the Commander observes. The General has written a very concise article that provides an excellent template for any Operational Level staff's use when training for both deployment and certification. The "best practices" discussed in the article certainly permit utilization of the UJTL as a basis for operational thought and employment, and allude to the need for simple quantitative measures to provide a basis for the complex and necessarily subjective assessments that are necessary to effectively employ forces and coherently communicate both vertically and horizontally[5] (18).

So, if standards, notably primarily numerical values, can be associated with tasks to assess performance, how can a warfighting scenario be assessed? The weighted

---

[5] Vertical and Horizontal communications are commonly used to describe communications with equivalent levels of command (in this case other Operational Level commanders, FCCs, (Horizontal) or to both senior and subordinate levels. (Vertical)

average methodology developed by McGinnis and Stone provides insights and is at least one method for proceeding. However, as noted, this simple summation technique may not be sophisticated enough. Initial research into potential alternative methods for approaching performance assessment led to work conducted by psychologists that addressed some of the issues with team performance that might have potential application.

The acquisition of skills and abilities, especially the types being assessed in this proposal, are almost exclusively cognitive in nature. It would seem appropriate to use a formulation representing cognitive skill acquisition as a potential method of performance assessment. The specific method of aggregating skills or measures would need to be determined, and it also appears that if multiple steps of aggregation were necessary, the additional complexity introduced might be too difficult to integrate into any cognitive methods. If, however, the aggregation could be accomplished in a single step, then perhaps a mathematical representation of skill or learning could be utilized. One mathematical equation that could be used was researched thoroughly by Newell and Rosenbloom and is shown in equation 2.2. This equation is known as the "power-law of practice" and reflects the fact that the time to accomplish a specific task decreases with an increasing number of trials (19) .

$$T = A + B(N + E)^{-\alpha} \tag{2.2}$$

where A = asymptote of learning (the task can be performed in a short time after enough practice)

B = Range of learning (scaling parameter)

N = trial number (number of practice sessions)

E = a function of the number of previous trials (prior experience)

α = a "learning rate" parameter (power law exponent)

and T = expected reaction time

A key aspect of the use of the power-law of practice is that the various examples that were discussed involve specific tasks related to motor skills, elementary decision making, or some problem solving tasks. Individuals performed each of specific tasks (19).

A follow-up paper presented at the International Conference on Complex Systems in 1998 by Mayer-Kress, Newell, and Liu discussed two classes of learning curves. One was an exponential form as depicted in equation 2.3 (20), and the other was the power-law as shown in equation 2.2.

$$p(n) = A + Be^{\gamma(n+n_0)} \tag{2.3}$$

Where: $p(n)$ = the performance function

A= a constant corresponding to the asymptotic performance parameter

B = the change from initial performance

γ = the magnitude of the learning rate (generally negative)

n = the number of practice sessions.

n_0 = a constant to account for a shift in the start time

The exponential form is used when there is a constant learning rate and a fixed time scale. Mayer-Kress noted that the use of the power-law of practice approach relied on learning rates (the exponent in equation 2.2) that were individual and task specific. They also hypothesized that it appeared possible to use the exponent, which might permit a more universal characterization, but that at the time that their article was published

there was very little empirical work to attempt this characterization. Mayer-Kress was able to show that with appropriately chosen exponents, the exponential form of the learning curve could approximate a power-law of practice, but again, the activities chosen for data calculation were simple tasks and not related to group activities.

Heathcote, Brown, and Mewhort conducted further research into the applicability of either power-law or exponential formulations to depict human learning. They concluded that the power-law development and use described by Newell and Rosenbloom was flawed because it was based upon averaged data. Heathcote found that if data from trials was not averaged, then an exponential function provided a better fit for trial data, whereas if trial data was averaged, then a power-law seemed to produce better results. The authors were concerned with the number of parameters used in the power-law formulation, since there was an additional term not used in the exponential form. They felt that the additional term made it possible that a general power function could mimic an exponential function, but the ability to compare the power law to a simple exponential form was nearly impossible (21). The overall conclusion of this research team was that an exponential function was the best candidate for a Law of Practice instead of a general power function.

The discussions above tend to make use of either of these formats increasingly less attractive. A final article was reviewed that discussed exponential and power law functions, with the hope that it might provide a usable format of a learning curve approach for this research. Ritter and Schooler reviewed both the power law and exponential forms of the learning curve. Their insight was that if learning followed an exponential, then learning was based on a fixed percentage of what needed to be learnt. Further, if learning followed a power law, then learning slows down in that it is based on an ever-decreasing percentage of what remains to be learnt (22). The authors concluded by saying that the learning curve was a success story for cognitive psychology, which had shown that learning is ubiquitous and has provided mathematical accounts of the rate. They further state that the learning curve is a success for cognitive modeling (22).

An intriguing approach for performance assessment is an alternative aggregation approach developed by Startin. The method utilizes a combination of weighted averaging and logarithmic functions to calculate performance assessment for individual warfare areas within a Navy Strike Group. The method retained the conditions and standards developed at the tactical level for _METLs, since those measures already assign values to many tasks. For the higher operational level, utilizing a combination of the standards that currently exist at various organizations would provide the basis for assigning values to the tasks required for performance assessment. The Startin approach groups measures such that they provide information about processes, outcomes and circumstances. The evaluation of processes, outcomes, and circumstances requires that _METLs have skill categories, process and outcome groupings, and finally circumstance groupings. He emphasizes that the fidelity of an assessment is dependent upon the quality and completeness of the set of _METL measures and the process for their evaluation (23).

A very distinctive feature of the method proposed by Startin is the use of logarithmic functions for characterizing staff performance to approximate the "learning curves" discussed by cognitive scientists. In the research necessary to develop a method to mathematically describe "the learning curve" of a staff (which can be geographically distributed, but serve a single commander) Startin developed criteria that would be advantageous or preferred if they could be met:

- The function should be continuous and monotonically increasing, which would ensure that the function is mathematically well defined and increased as the input variable (skill level, S) increases,
- The function value should diminish more rapidly as the input parameter (S) decreases,
- Function value should increase more slowly as the input parameter (S) increases. Preferences 2 and 3 ensure that the behavior of the function is consistent with data concerning learning curve patterns. That is, improvement with practice occurs rapidly at the beginning (when performance overall is poor), but the pace of improvement with practice slows as performance improves,

- Insensitivity to small changes in input parameter, which ensures that the value of the function does not unexpectedly vary between small or large input values,

- The product of several driver functions is not reducible to a simpler expression, which will ensure that critical skills will have independent impact on the overall assessment score, and the assessment will not dilute the impact of critical skills on the final score (23) .

All the characteristics described above are desirable; they include the features that are inherent in the power-law formulations or the exponential variations (2 and 3 above) as well as provide other desirable traits. The insensitivity to small changes to input parameter is useful since variances, which can often be large, are not unusual after a significant period between evolutions or as organizations become familiar with tasks. Finally, the trait describing the product of several driver functions is key in that it is very useful to senior leaders to be able to see an overall score, but also see where the impact of either strengths or deficiencies of critical skills are affecting the staff's progression. This trait is especially desirable for ROI discussions, since it will, when coupled with the traits 4 and 5, portray a staff's "steady state" performance and permit examination of methods to improve specific organizational element weaknesses. The method he proposes for assessment is to average the scores for any specified set of skills. This is acceptable since those skills relate to one another and will not be averaged with dissimilar skills. Skills are then assigned weights, multiplied by the appropriate metric, and summed. The overall weights of all associated skills are also summed and used as a divisor to normalize the skills in that grouping. The averaging method to be used varies slightly from the one developed by McGinnis and Stone in equation 2.1, but it may well lend itself to an expansion of this particular method as well as the McGinnis and Stone method, and further discussion will occur later in this dissertation. Since the measures have also acquired weights according to the processes where they were assigned to, these weights will also be accounted for in the averaging (23).

$$S_i(T_j) = \frac{\sum w_j T_j}{\sum w_k} \tag{2.4}$$

where S_i = i^{th} pillar, applicable to the execution of a specific organizational pillar

T_j = j^{th} metric in the i^{th} JMET-process bin, representing observed measure data

w_j = factor between 0 and 1 representing the relevance of the metric T_j to pillar S_i

This methodology is designed to reflect skill proficiency and not task capability as is currently observed in PBViews. The method is similar to the McGinnis and Stone method since weights can be assigned based upon the relative importance of each task.

The methodology then requires aggregation to reflect overall performance assessment. Each skill, whether it be process, outcome, or circumstance has been evaluated, but these must then be combined to assess overall performance. Startin proposed using driving functions (composed of the weighted average skill areas in equation 2.4) relevant to specific warfare areas that relate to either the ability to learn or the half-life of skills as a means of approximating learning curve behavior. These driving functions tend to be logarithmic, and thus are insensitive to small changes and more accurately reflect skill increases or decreases. He then proposes normalizing the driving functions so that no individual warfare area has a value greater than unity, and thus the normalized functions will be in a range of 0 to 1. The final step is to take the product of all the driver functions. The number of driver functions multiplied is dependent upon the number of skill categories measured (23). An example of this concept applied to this research would be to view an area such as Logistics for an Operational Level staff. The assessment of this area relies on viewing nine specific skill areas. Each would have a driving function and normalizing function (which is developed in subsequent paragraphs) which are multiplied by each other to arrive at an overall performance assessment for the Logistics Organizational Element

One concern in the assessment process is that many events and indeed processes occur in parallel and though this is completely normal, it is difficult to assess performance in a system that is not sequential. The weighting average method proposed by McGinnis and Stone could be taken to mean that events must occur sequentially, in order to be "rolled up" to the next level. That might not mean that parallel subtasks or tasks had to be performed in any given sequence, so it appears that the weighted averaging method would work in the Operational level staff application. There is no doubt that the daily activities of any type of staff are not only performed in parallel, personnel who are already involved in a completely separate task may also perform the two or three other tasks. Another issue that might impact a weighted averaging method, or any method, is the fact that it may not be possible to assign a single _MET to a single function within a staff. A cursory review of the measures used to assess _MET performance appear to permit assigning individual measures to some sorts of sub-groupings to permit performance of functions, and that will be the key to the aggregation strategy used for this proposal.

Startin (23) investigated the possibility of aggregating measures based upon the preceding discussion and proposed that a combination of weighting and averaging and some product function be utilized to arrive at a more accurate depiction of training readiness. He chose a logarithmic function as the means to create factors for a product relationship because a logarithmic function obeyed the guidelines set forth in the preceding paragraphs, especially the concerns with accurately portraying learning (as proficiency increases, the rate of improvement decreases) as well as the insensitivity to single values that might change dramatically in a short (day-to-day) observation.

The major feature of this method is the use of Logarithmic Driving Functions (LDF), which are so termed because they will drive the assessment scores. The LDF utilizes the skill level values as an input. The output of the driving function is the value that reflects the impact of that skill on the overall readiness of the specified staff area. The overall score for either a specified staff organization or perhaps even major functional area is a product of the LDFs relevant to that organization or functional area.

The original development concerned specific warfare areas, but the concept for this proposal is to treat either staff organizations or major functions as warfare areas were treated in the original work. The principal new development that Startin was able to introduce into the assessment process was the specific LDF that would be able to utilize weighting and averaging, and also meet the conditions stated earlier in this discussion regarding desirability. His investigation led to an equation in the following form (23):

$$LDF_i(S_i W_i) = \log_{10}\{1 + 9[S_i + (1 - w_i)]\} \qquad (2.5)$$

where S_i = Score for i^{th} skill(S_i is a real number between 0 and 1)

w_i = Skill impact parameter (weight) (a real number between 0 and 1)[6]

Startin noted that it would be possible for LDF_i ($S_i w_i$) to assume a value greater than 1 when w<1, so that a product of LDFs might result in a value of greater than 1 as well which would have no meaning. (The assumption being that the upper range of score would be 100%, not a quartile system as in PBViews ranging up to 200%) To ameliorate this issue, it was necessary to devise normalizing factors for the LDFs.

As already noted, each weighting factor w_i has an impact on the maximum value of the LDF when S_i = 1. The desire would be for a critical skill that was performed correctly to have a LDF_i = 1. This desire was the basis for Startin formulating a normalization factor in the following form (23):

---

[6] A note to the above is that a w = 1 would signify a high impact skill, and if S=0, then the LDF =0. If w=0, and the LDF =1 even if S=0.

$$A_i = \frac{1}{LDF(S=1, w_i)} = \frac{1}{\log_{10}(19 - 9w_i)} \tag{2.6}$$

Having developed a method to arrive at a value for a skill, to calculate LDFs, and to normalize them, the remaining task was to aggregate the LDFs in a manner that would represent the performance score for any warfare area. The process for this proposal will be to aggregate the LDFs to produce an assessment score for either an organization or for a major functional area. As discussed earlier, the most logical way to combine the LDFs seemed to be a product relationship, which would represent the parallel accomplishment of tasks. Further, if a specific measure was not accomplished at the time of the assessment, then that measure is simply replaced by a unity factor, and the measure with its appropriate weight can be included in a later calculation.

The product function of the LDFs that also uses the normalization factors in equation 2.4 takes the following forms (23):

$$P = B_K \prod_{n=1}^{K} \log_{10}[1 + 9(S_n) + 9(1 - w_n)] \tag{2.7}$$

and

$$B_K = \prod_{i=1}^{K} A_i \tag{2.8}$$

where S_n = the n^{th} skill-category score

w_n = n^{th} skill impact parameter

B_K = the product of the LDF normalization factors

Equations 2.4 through 2.8 will be used in the methodology section as well as equation 2.1 for the calculations for performance assessment values. Though more complex than the methodology proposed by McGinnis and Stone, anecdotal results from the use of this method were very encouraging but not pursued.

Tillson suggested that commanders at all levels required a tool that could enhance their ability for reporting readiness that could serve a three-fold purpose. First, to provide a framework that could identify the tasks essential to performance of joint missions, secondly that the tool should provide a means to communicate the chosen tasks vertically to assure common understanding, and finally the tool would provide a way to report the joint mission-oriented readiness of assigned forces (24). He states three underlying principles, which supported the development of such a tool:

- The COCOM Joint Mission Essential Task List (JMETL) and Service METL should be connected to provide two-way mission related information flow between the COCOMs and the Operational forces,
- The Services should adopt a common measure of training readiness that was both mission and task related. This would assure that the COCOMs would be able to clearly understand training readiness on a common, mission-oriented basis,
- The COCOMs and Services could employ modern data base management systems and the Global Command and Control System to integrate Service and Joint data bases in a manner that will place responsibility for maintaining the data bases with the appropriate COCOM or Service (24).

Eleven years later, the implementation of the Defense Readiness Reporting System (DRRS) (25) is beginning to approach the conceptual proposal of Tillson, but it is still not fully implemented or understood. Tillson did not advocate that Services abandon their reporting systems, but that they be given a clear understanding from the COCOMs regarding the tasks required at the Strategic and Operational level so that the Services could develop Service specific METLs at both the Operational and Tactical levels that could permit the information flow discussed above. Each Service has its own version of

METLs, but until the Operational Level JMETs are commonly used for training scenarios, training and readiness reporting, and resource allocation, there is little urgency for the Services to develop appropriate METLs for themselves.

Tillson also acknowledges that even with the lack of an overarching JMETL (which is being developed for use in DRRS as this research effort is being undertaken) the hurdles to his second and third points are not trivial. He illustrates training readiness in terms of U.S. Marine Corps aviation, and notes that the U.S. Navy system is similar and mission based. Unfortunately, even though all four branches have aviation readiness systems, there are myriad other mission areas that are combat specific depending upon theater and geography that would require significant innovation to become standard, but perhaps the answer isn't standardizing these measures, the correct approach may be adopting METLs to the missions required, which certainly is possible with the structure of the UJTL.

As briefly mentioned earlier in this survey, a study recently conducted by the Army Research Institute (ARI) investigated, "employing structured simulation-based exercises to explore and assess the methods, tools, and measures necessary to facilitate distributed wargaming. Key aspects of the environment were structured wargaming exercises for distributed staff and command groups, collaborative tools, and a set of performance measures to assess wargaming (14)." The environment studied had many similarities to the architecture, scenario basis, and objectives of the Navy's FST program. Further, the research addressed objectives, which are essential to the success of the FST program as an operational training capability, such as the need for a network supporting information displays, and communications systems representing the full spectrum of capabilities used in operational theaters. Two other specific objectives that are in concert with the requirements necessary to evaluate, assess, and certify a CJTF or FCC staff are:

- Design and develop structured exercises to support the conduct of planning and wargaming in a simultaneous, collective, multi-echelon and distributed manner. The distributed exercises require planning between higher and lower echelons (vertical integration) and across the same echelon level (horizontal

integration). Tailor the design to an audience in which three to eight participants located in at least three separate or distributed locations interact directly and accomplish tasks collectively. Include collaborative, interdependent tasks (i.e., the task requirements for each participant will depend on the work of the other participants).

- Design and develop measures to assess the effectiveness of distributed wargaming. Performance assessment must address the outcomes of distributed wargaming to identify whether participants successfully identify and address problems or conflicts within a COA. Conduct a formative evaluation to gather participant feedback on the research environment, particularly the wargaming exercises and measures developed (14).

A challenge faced by the authors of the report was the fact that previous synthetic exercises had not utilized extensive horizontal or vertical elements. Though the horizontal and vertical elements were not necessarily viewed at the Operational Level, the principle is valid and is a key consideration in assessing performance, cost, and ROI at the FCC and CJTF staff level. The observations of the participants in the investigation did not find Horizontal engagement to be particularly difficult, but found vertical engagement to be a challenge. At the operational level, horizontal engagement may indeed prove challenging since distributed simulations typically utilize on-site representatives with pre-scripted responses to portray the reactions of another FCC. In the ultimate configuration, a distributed simulation would permit actual FCC staffs to interact with each other as well as with the CJTF staff replicating the same architecture and capability seen in either live training or real-world operations. The observations of the participants in the ARI investigation are germane for both interactions at the operational level. The observations, which require incorporation into the current project, are:

- Participants in the vertical condition reported that rarely would subordinate commanders have the opportunity to participate in wargaming,

- Participants in the Vertical condition stated commander-centered wargaming seemed more like mission rehearsal,

- Individual differences in wargaming experience undoubtedly affect performance. Future research might use background information, from demographic surveys, for example, to assign participants to duty positions and roles (14).

The summary of the ARI report noted that there was a need for more research and training permitting the Army to train in a commander-centered wargame across echelons to meet its Future Force objectives distributed operations (14). The synthetic training carried out by the Navy, Air Force, and under development by the Army permits all three Services to train with or at least interface with FCC staffs, thereby enhancing and improving the relevance of synthetic training. This capability cannot be truly viable until a method to assess the effectiveness of the staffs involved, the costs involved, and the ROI gained from the training can be reliably reported. Certainly, the Services derive no long-term individual benefit by supporting such training, even if it could be supported without jeopardizing individual, small unit, or deployable unit training objectives.

A discussion regarding the possible extension of performance assessment methodologies developed in the course of this research proposal to non-military applications addressed several organizations, including commercial utilities. Commercial aviation, and now the commercial nuclear power industry are becoming increasingly dependent upon the use of simulation and performance assessment in complex casualty control scenarios as normal practices. Spurgin discussed the possible uses of data collected during simulated accident scenarios for the purpose of assessing training effectiveness. He noted that if operator actions, plant responses, and other observational data from training scenarios was analyzed that insights into performance strengths and weaknesses, as well as the strengths and weaknesses of the underlying training program could be determined (26), (27). This observation is a commercial application of the principles set forth by the founder of the U.S. Navy Nuclear Power program, Admiral Rickover. Performance assessment, direct feedback, and training, which addressed shortcomings, observed in both the level of knowledge of the operators as well as their operational proficiency was and remains one of the bedrock principles imbued into U.S. Navy nuclear trained operators. His philosophy of training and education has been a model for operational effectiveness for over 50 years (28). The major difference between the training proscribed by Admiral Rickover and that becoming commonplace today was the use of simulation. In the 1950's and until the late 1990's simulation was unable to

properly model the conditions and consequences of operator actions. As such, all training was conducted on actual power plants so that there would be no doubt as to the results of any action. None-the-less, the use of very specific performance standards and concept of team performance evaluations has been used as a foundation for performance assessment both inside and outside the military

Current simulation and stimulation capabilities, especially in a distributed environment have eliminated the concerns Rickover had regarding realistic operational conditions. Regardless, his overriding desire to train in the most operationally realistic and challenging environment possible, to continually assess the performance of both the personnel and equipment, and to improve processes, equipment, and procedures are existing examples of the concepts of evaluation, cost accounting, and ROI advocated by the various authors who have advised the industry and government institutions cited above.

The ability for simulation, especially by federating simulation systems from all the Services, to foster and enhance Joint training has been recognized. A 2004 article in National Defense Magazine (NDM) (29)pointed out that the development of the Joint National Training Capability (JNTC) was beginning to permit the Services to connect simulations, which permitted operators to participate in training exercises with robust adversaries, and allied forces without having to be concerned whether these forces were live, virtual, or constructive. Since that article, the JNTC has expanded and each Service has continued to develop methods to federate simulation and stimulation systems into the architecture. The issue is no longer one of being able to prove connectivity, but of utilizing the system to accomplish training and evaluating the performance of the participants. The network can support operators on consoles or in vehicles as well as CJTF or FCC staffs coordinating multiple Services or components in a very geographically dispersed theater.

The need to improve combined operations and the most promising way to do so was identified as early as 2003 by Joint Forces Command. Harold Kennedy cited efforts by Admiral Giambastiani, the Commander of Joint Forces Command "to assist unified

combatant commanders find better ways to organize multi-service units (30)." These efforts were largely focused on developing a Joint National Training Center, which would permit US and eventually Allied forces to train together. Further, the need for Standing Joint Force Headquarters (SFHQ) was also identified. As stated in the NDM article,

- "The SJFHQ is a team of operational planners and information command-and-control specialists, Bartell said. When the combatant commander decides to launch a JTF, the SJFHQ's job is to help make it happen with a minimum of confusion and delays. When a JTF is established, the SJFHQ takes over as its headquarters staff, enabling more proactive and coherent planning and quicker use of capabilities than can be accomplished by the traditional ad hoc approach, Bartell said (30)."

Each Combatant Commander was to have an embedded SFHQ within its staff by 2005. Thus, the need to be able to train and assess both the FCC and CJTF staffs with reliable, repeatable, and useful measures is necessary to facilitate each SJFHQ's ability to quickly deploy and commence operations.

2.1.2 Issues with Current Assessment Methodologies Requiring Resolution

The essential premise of the performance assessment methodology is that there is a method of combining discrete measurements that result in an accurate assessment of either overall performance or specifically selected functions. Further, it should be possible to view performance from the start of training to its completion with discrete steps (typically, daily evaluations) which reflect changes (either an improvement or decline in performance) in a manner which represents a "warfighter's" view. The ability to aggregate measures and then perform calculations, which portray performance in relation to accepted standards must be clearly defined and accepted by the COCOMs, Services, and Operational level Commanders for this proposal to be useful. The three possible methods for aggregation that the researcher believes will permit effective performance assessment and have potential for adaptation to this research are some form of weighted averaging, the application of learning curve power-laws, and an aggregation

and logarithmic function approach that might have potential in achieving the research goal of this proposal. Each of these will be addressed in subsequent paragraphs, and then referenced where applicable in the methodology development.

The mathematical formulations presented in association with either the power law of practice (Equation 2.2) presented by Newell and Rosenbloom (19) or the exponential form (Equation 2.3) developed by Mayer-Kress, Newell, and Liu (20) appear to have merit for the collective learning environment that is being investigated here. Unfortunately, though the mathematical formulations above have some attractive features, there are some issues that make using them for evaluation at the operational level difficult, if not impractical. All the results above rely upon very large numbers of trials, the power law is defined in terms of a time to become proficient, both equations require an initial skill capability, and finally, not only are these expressions applied to simple tasks (even stacking cards is simple next to analysis of an enemy's plan) they also involve individuals.

The skills referred to in the above discussions are typically motor skills, or in some cases a mix of a cognitive skill (e.g. selecting the proper card) and a motor skill. The skills necessary for successful planning, execution, or assessment at the operational level are completely different. Indeed, referring to the traits necessary for success at this level may be a misnomer, but as long as it is understood that skills refer to complex communications, problem solving, and decision making abilities, vice very elementary motor skills, then confusion should be reduced. An area that is also almost impossible to quantify for the types of groups being assessed at this level is experience. Thus, some of the initial conditions would require quantification that is vastly different from the number of times a specific motor skill has been employed.

Further complications arise when attempting to characterize diverse groups, which perform some tasks very often while the same group performs others (for example working in the area of foreign humanitarian assistance vs. some sort of combat operations) infrequently may prove very difficult to quantify. Further, the group would have to be "averaged" somehow for range of learning, as well as characterizing the same

group with a single "learning parameter." Some cognitive scientists have also expressed disagreement with the application of the power-law model, describing it as an "artifact" because as described within the staff above, it is not really accurate or proper to average the varying learning rates of different people (21).

The final concern with the power-law approach is that the quantity arrived at through the calculation shown is in terms of time. While useful it really does not result in the desired quantity, a performance assessment value. Though it might be possible to create a relationship that used the time to reach a point of diminishing returns (in this case, the point at which learning rate stopped changing) and then arrive at a value for the corresponding skill level, there would be a danger very similar to the arguments against averaging averages in the previous discussion. If learning rates, ranges of learning, and numbers of trials varied greatly, then either some distribution which might possibly miss the impact of a new or overly experienced member would have to be used, or, to minimize impacts, an "average" team member would have to be constructed. The last resort would be to only use the specifics of a selected individual, most likely a group leader to arrive at the resolution time quantity. All of these are complex, even more suspect in terms of method, and thus ultimately, though interesting, use of this particular approach is not suitable.

The attractiveness of either the power-law of practice or an exponential equation to assess performance lacks sufficient basis in a group context to be practical for this research, so some other mathematical formulation is necessary to support performance assessment.

The current system being used by the U.S. Navy to evaluate Strike Group training readiness is Professional Business Views (PBViews), which is a product of PerformanceSoft of Ontario, Canada. This tool utilizes a color coded hierarchical display capability that provides the opportunity to view areas that are significantly deficient contributing to a related higher-level deficiency. The methodology utilized by PBViews is an averaging and weighting scheme, which, at first glance, appears very similar to the method proposed by McGinnis and Stone.

Dr. David Rodney conducted an analysis of the ability of PBViews to accurately aggregate weighted and averaged measures from the NMETL to accurately reflect readiness in 2003. One of the conclusions of his research was that the weighted averaging scheme used by PBViews was inherently inaccurate (31). He found that it was possible for a series of metrics associated with a specific mission area to be evaluated as T-4 (unsatisfactory), yet if other metrics within the same parent area were satisfactory, the parent area would be satisfactory. The reason for this finding was that if all measures are "critical" and weighted accordingly, the process becomes a simple averaging of all the measures, thus if a preponderance of measures are satisfactory or better, then it is quite possible for even a subtask area to be satisfactory, even though it may have significant deficiencies. Three years of observation by the researcher bear out, at least anecdotally, this finding. The results of data collected from over 50 Strike Group evaluations resulted in only occasional mission area scores below a satisfactory level, despite significant deficiencies requiring immediate remediation prior to Strike Group deployment.

There are several reasons why these results can occur. The predominant reasons are the selection of measures to be used, the inherent characteristics of PBViews, and the scoring conventions applied to binary measures.

PBViews is based upon a linear scoring convention that divides performance into quartiles (also known as "index values) ranging from 0 to 200%. Satisfactory scores range from 100 – 150%, with two subordinate levels, one of which is unsatisfactory, ranging from 0 – 50%. The observer compares performance of a specific task to the assigned standard thereby determining an index value. The observed value is then assigned an interpolated numerical index based upon the quartile scoring convention.

Strike Groups are evaluated in terms of Warfare Areas, which are delineated by the NMETL. The Warfare Areas correlate to Parent Tasks or mission areas from the McGinnis and Stone article. There are specific Navy Tactical Tasks (NTAs) for each NMETL, as well as subordinate tasks for the NTAs. PBViews computes the values for each NTA as a weighted average of the subordinate measures for that NTA. The overall

score for each Warfare Area is then computed as a weighted average of the subordinate NTAs. The overall Strike Group score is then computed as a weighted average of the Warfare Area values. Again, the scheme detailed here would seem to be in concert with and actually a validation of the method proposed by McGinnis and Stone.

The departure from the methodology of McGinnis and Stone is that the default behavior of PBViews weights each Warfare Area, NTA, and subordinate measure equally at each level. Obviously, as a Warfare area or NTA has increasing numbers of subordinate levels, if each is equally weighted, then it is entirely possible for a significantly deficient grade to be either elevated, or more typically diminished in importance merely by the application of the default weighting scheme. The overall result of this scheme is that measures become weighted by quantity rather than by importance. The default weighting scheme can be tailored, which does provide a more realistic overall assessment, but the issue then becomes, is the weighting scheme approved by one Commander for a given scenario acceptable to another?

The issue, especially in a military context, is how to relate these two approaches into training evaluation. The phases of NMETL[7] development and utilization share or can be directly linked to both Sloman's interactive ISD and the six step methodology proposed by McGinnis and Stone (8). The four phases necessary to assess military capabilities are: mission requirements, planning, execution, and assessment. Properly applied _METL analysis permits assessment (in this case the assessment phase of the capabilities phase is the feedback phase discussed by McGinnis and Stone) of each of these phases by applying a task, a condition, and a standard to each (32). McGinnis and Stone break the measurement of effectiveness into tasks (subtasks are equivalent to subordinate _METLs for this discussion), conditions and standards as well. The major difference in the two treatments is the discussion of performance outcomes very often being assessed as pass/fail (8).

---

[7] In this treatment JMETL or NMETL are an interchangeable terms, so _METL will be used in subsequent references to either a Service specific or JMETL when a general application is required.

At the individual level, since many tasks can be evaluated in a pass/fail manner, and weighting is rarely an issue, the impact of a binary assessment is not significant, and is useful because the skills requiring mastery really must be mastered before the individual can be integrated into a team. McGinnis and Stone, as well as Kirkpatrick note that assessing team performance is more difficult, and in many cases, a binary approach either inappropriately raises or lowers overall assessment. The impact of a binary approach is just beginning to be understood in Navy Strike Group evaluation. The current evaluation system, Panorama Business Views (PBViews) is an industrial application that has been adapted for use in evaluating Strike Group performance (which is equivalent to an Army/Marine Corps Division or Air Force Air Wing in size and command organizational structure). The software was designed to evaluate (33) manufacturing performance and has only recently become available for other applications. The Navy has been using this application since 1999 and has worked to tailor the software (within license restrictions) to accurately assess NMETL accomplishment. The ability to look at sub processes and quickly generate "snap shot" reports makes this software application attractive, but the algorithms used to calculate performance are not well understood. The default behavior of PBViews is to weight each warfare area, functional area, tactical task, and metric equally at each level. This approach is quite different from the approach proposed by McGinnis and Stone which weights tasks and subtasks at each level according to their particular merit. The unintended PBViews result is that metrics are weighted by quantity rather than importance (23). The situation is exacerbated by the use of binary measures since, in an area where there might only be five metrics, a single fail (0) will require four passes (100) grades to achieve a rating of 80 (23). If the other metrics are standards, then, the average performance will make it nearly impossible to assess a particular area as "passing." A final issue with this sort of application for assessment is that PBViews aggregates scores by averaging all scores in any given area. Thus assigning weights as proposed by McGinnis and Stone and as is mandated by the _METL assessment process can be superfluous. If one completely understands the algorithms involved in assessment, then it may be appropriate, in very limited instances to assign binary grades,

but, especially in a war fighting context, and in a team assessment, it would be more appropriate and insightful to assign quantities to measures in every possible case as proposed by Phillips (16). Intangibles can be measured, planning is an activity and it has definite results as Cokins observed. The vast majority of activity pursued at the senior levels of corporations, in the military, or in emergency response is team oriented and involves planning and decision-making (15).

It becomes apparent over the course of time that independent of the structure for any operational template for Warfare Areas, NTAs, and subtasks, the final score is actually no more than an average of all the scores combined. It is the natural consequence of averaging averages. Though weights are assigned for even individual measures or NTAs, the effective weight will be the product of the weights assigned for the measure, the NTA, and the Warfare Area. This abrogates the original intention for assigning weights. It is possible to have only a single weight for a measure, which then becomes the weight for the NTA, or there may be multiple levels of weights as mentioned. In the aforementioned case, it might be possible for a single measure to be more "important" than the measure in a series of measures, NTAs, and a Warfare Area deemed critical for mission success.

The last area, which is of concern in any scheme, but once again tends to be exaggerated in the PBViews scheme, is the weighting of "binary" measures. Typically these are assigned either 100% or 0%. From the preceding discussion, it is clear that either score can unrealistically bias either a subtask or NTA. This tendency has been recognized and corrected. The convention adopted by the U.S. Navy is for a Satisfactory (Yes) binary measure to be considered Satisfactory, and an Unsatisfactory one (No) to be considered Less than Satisfactory, but NOT Unsatisfactory. (The result is assigning a value of 75% for the 50-100% quartile in PBViews vice a 25% or O% from the lowest quartile.) This same approach is going to be used, with some modification for the assessment methodology approach in the final section of this proposal.

2.1.3 Proposed Performance Assessment Solutions

If a system like PBViews is not the answer, what solution can be pursued that realistically addresses collective performance? Regardless of the approach, there should be an ability to address increasingly but a more robust and sophisticated method (as suggested by McGinnis and Stone) is required, how can the approach which accounts for conditions, standards, and tasks, whether _METL or not be implemented in the real world? Regardless of the methodology ultimately used, a necessary assumption is that all tasks can be assigned representative conditions and standards. These conditions and standards would necessarily have to be agreed upon by a mixture of personnel in whatever organization is assessing performance.

Based upon the foregoing discussion, there does appear to be at least some merit in attempting an averaging approach to scoring at the Operational Level. The problematic averaging and weighting issues involved with PBViews must be avoided, which will demand significant effort in determining and assigning weights for tasks and subtasks. It is also very important that those areas deemed most important do not become lost in an averaging of averages, or product of weights aggregation. The use of an averaging approach will be further discussed in the methodology portion of this dissertation.

The currently employed methods for evaluating Operational Level staffs rely heavily on qualitative methods. These methods are applicable to evaluating the commander, as described in General Gary Luck's article "Insights on Joint Operations: The Art and Science (18)." The evaluation of the supporting staff processes and products can be approached quantitatively, and ensuring that the quality of the inputs to the Commander are the best possible, may not guarantee the best possible decisions in every case, but without credible input, the decision making capability of the Commander may be hampered by far more than the uncertainties which are a part of the "fog of war."[8]

---

[8] Though commonly applied to combat operations, there is also a degree of uncertainty in all operations across the entire Range of Military Operations (ROMO).

In the absolute sense, there is only one method to evaluate a military organization's performance and the ROI achieved. The starkest case would be after Major Combat Operations (MCO) where the only metrics that apply are whether or not the conflict was won, the numbers of U.S. casualties, and the final disposition of the adversary (34). This is not a practical method of assessing performance, costs, or ROI, even in an environment that is not resource constrained. The only way to evaluate performance, which translates directly into readiness, is to do so prior to the employment of forces, and to incorporate all that has been learned from similar operations into the preparation for the organization that will deploy next.

The current operational tempo for the U.S military, including the Operational Level staffs is such that preparation methods, which were acceptable as recently as 10 years ago, cannot address the needs of the units to be trained. Though synthetic training has become commonplace for individuals and even some unit and force level applications, there is a need to validate this training in some live venue. That is emphatically not the case with operational level staffs. Since the primary functions of these staffs are to develop plans, issue orders[9], and assess the effects and impacts of those orders, and since at the Operational Level, very little of the activity centers around actual employment of weapons or systems, which significantly decreases the reason to train these types of staffs in a live environment. The issues that must be dealt with by the Commander are best presented in a synthetic environment since they are largely political or concern large geographic areas. Thus, though live environments afford some realism that cannot be replicated, the areas requiring focus for an operational level staff is probably best served in a synthetic environment. This environment provides equal and in some cases superior opportunity to observe reaction to events that are not possible in any other arena, thus there is also the opportunity to measure the responses in a quantitative manner. The current evaluation system certainly recognizes these opportunities, but there

---

[9] Issue orders will be used to denote both transmission of an order, or executing it, and the subsequent monitoring of the order which are the primary functions of the Operations Center watch team.

is still not a uniform methodology for measurement, which is a deficiency, deserves attention (18). What methods or documentation are available to ameliorate this problem? One of the basic documents is the Master Training Guide (MTG) (35). This document, when coupled with known training objectives does permit evaluation of performance, but the basis for the evaluation tends to be primarily binary in nature, that is, either an action is completed or not. There are two issues with an evaluation of this sort: 1) It is possible to assess that there is a satisfactory level of performance, but little else. 2) The MTG is very extensive, and it is unrealistic to expect that any evaluation team or even staff in a self-evaluation mode would be able to complete all the sections of the guide, and certainly not in a repeated fashion or in a limited time frame.

Even more importantly, there are key tasks which are repeated often and that require demonstrated staff proficiency. Further, at the Operational Level, the Service origin of the staff may be unimportant. If the staff is Serving in a Functional Component role, this would be important, but in operations where any operational level staff may be employed across a Range of Military Operations, including very specific missions such as such as Foreign Humanitarian Assistance, Consequence Management, or Civil Support (2) than any JTF staff performs the same tasks, the Service origin only providing a background for perspective.

The results of equation 2.1 rely heavily on appropriately choosing the weights, that is, the importance of each subtask or family of subtasks. The methodology proposed by McGinnis and Stone relies on weights being in a range between 0 and 1. The authors also note that weights could be derived mathematically or elicited from unit personnel. The attractiveness of this method is that the evaluation of almost any military organization, certainly an Operational Level staff, will necessarily have a similar construct in that the JMETs have measures. Whether the JMETs are grouped together and then averaged or the measures are decoupled from the JMETL and then combined in a different manner still requires weights to be assigned and combination at successively higher levels until a final scored in determined. A later section of this proposal will demonstrate an alternative method for using weighting which could also be applied to

performance assessment. The averaging methodology proposed by McGinnis and Stone is attractive in that it is very straightforward and for one or two levels seems to provide a method to arrive at appropriate weighting for each subtask grouping. Since the standards should be pre-defined, the only issue should be the appropriate selection of weights at each level.

The Startin method is one possible method that has potential for use in assessing performance in almost any Operational level training event, whether live or simulated. It could also be adapted, as discussed by McGinnis and Stone to be used in other settings such as Emergency Response scenarios at the federal, state, or local level since the skill sets required for each of these teams could be adapted to those in a military context.

McGinnis and Stone stress several important factors that were identified by Williams et al. First, the simulation training must be linked to strategic objectives of the organization. Second, it must also be directly related to the functions and processes of the organization, Lastly, training system development must demonstrate relevancy to real-world tasks in the eyes of the target audience, in this case the operational level staffs (36). When Williams et al. prepared their paper, although war gaming was conducted, simulation, especially in a distributed environment permitting warfighters to use emulators or actual equipment was rare (36). As related by Jean, current Navy simulations, Fleet Synthetic Training (FST), permit participation by USN, USA, USAF, FGN, and RN units and staffs up to a Component Commander level. The USAF exercises named Virtual Flag permit Air Wing participation as well as some USA ground and air defense forces, limited USN assets, and include the Component Commander. The scenarios imposed upon the units utilize representative threat capabilities and tactical employment. The ability to meet all three criteria stressed by Williams et al. is no longer a goal, it is a reality; sadly the ability to assess the performance of all units involved as appropriate, especially at the Component Commander level has not kept pace.

Sloman's discussion of e-learning and e-training emphasizes the fact that learning is for individuals while training is for teams (10). A system, which permits individuals reporting to staffs to gain the requisite knowledge to integrate into a team whether it is a

routine watch standing function, a planning team, or a support function, is vital. Distance learning affords such a capability, and with embedded evaluation tools, affords senior leader visibility on strengths, weaknesses and remediation. If the scoring tools are either relatable to or _METL based, then a given team's initial capabilities should be known.

One final issue for implementation of any of the above methodologies, including the basic method presented by McGinnis and Stone, is collection of data. Even though scenarios can be made to reflect real-world operational tempo and complexity, in multiple theaters of operation, the data collection requirement is often ignored or assumed away. Evaluators, especially at the higher operational levels, need to observe the performance of tasks and evaluate these in a warfighting context. The ability to provide accurate and relevant feedback is critical to the training process, and credibility as well as fidelity is necessary if the feedback is meant to provide instruction or correction. Many of the metrics that form the background for decisions are merely summations of events, yet a great deal of evaluation time is devoted to ensuring events are correctly tabulated. The final factor in the performance assessment equation is providing a data collection methodology, which frees evaluators to assess performance and not merely enter data. This is true regardless of organization, regardless of whether the training is live or a simulation, but is especially relevant for the military.

The final problem area, which he noted as perhaps the most complex, was the implementation of a Joint Training Readiness Reporting System. The passage of time and the ability to utilize databases or rapidly develop appropriate data bases, as well as the fact that since 2001, U.S. forces have been fighting in both a Joint and Allied environment should have at least diminished the concerns stated by Tillson concerning data base access, the need for COCOMs and Services to have access at the necessary levels to identify units, and the Service's ability to access any COCOM database in order to identify the conditions and standards of mission performance necessary for the requested forces (24).

There is discussion in previous paragraphs of this proposal regarding the value of assessments of training and readiness. Tillson recognizes the cultural stigma associated

with exclusively relying on external assessments to report training readiness, especially with the potential impacts on Commander's careers. He advocates using standards for assessing training, but for validating doctrine and training concepts as well. Finally, he suggests that though external evaluations play a part in assuring unit readiness is fairly reported, the ultimate reporting official and users of the standards should be the Commander's themselves.

All of the above is still as valid as the discussion presented by McGinnis and Stone. Regardless of the reporting system ultimately utilized, the focus, especially at the Operational Level, should be on the processes and skills of the staff of the Commander, not the Commander. Further, the ultimate authority for reporting readiness, even with an outside agency present for certification or re-certification is still the Commander, and thus the any reporting system and standards associated with it should be routinely used so that readiness reporting is consistent, regardless of the agency observing any training event.

There are methods in existence at the individual, single unit, or even composite organization (Carrier Strike Group (CSG), Expeditionary Strike Group (ESG), or Expeditionary Action Group (EAG)) to assess performance. The basis for some of the assessments performed for these organizations is some sort of METL theory, which could become at least part of the basis for performance assessment at the CJTF or FCC staff level. It is also very possible, indeed likely that existing assessment methods may need to evolve substantially or be completely revamped to be truly applicable, which should not be a deterrent to moving forward since this approach is still very much in the spirit of the efforts championed by the very distinguished authors cited above.

Since 2004, the US Navy has been expanding its efforts to utilize existing system training devices in a distributed environment. As recently as July 2006, successful exercises have been conducted by distributed US, British, and German ships or trainers, Navy and Air Force aircraft simulators, Army Air Defense simulators, and some limited ground forces. An article in National Industrial Magazine about the Fleet Synthetic Training (FST) (13) and the growing acceptance of this training capability concluded

with the statement that though training is just beginning at the Battle Force FCC level, it will soon permit the inclusion of other Joint partners and expanded Allied participation. There are assessments of staff performance conducted in the FST venue, but these require expansion and changed to realistically evaluate a FCC performance. The framework is available and will be investigated as a possible starting point for a CJTF or FCC assessment tool as well as using the cost data that is generated for this exercise series.

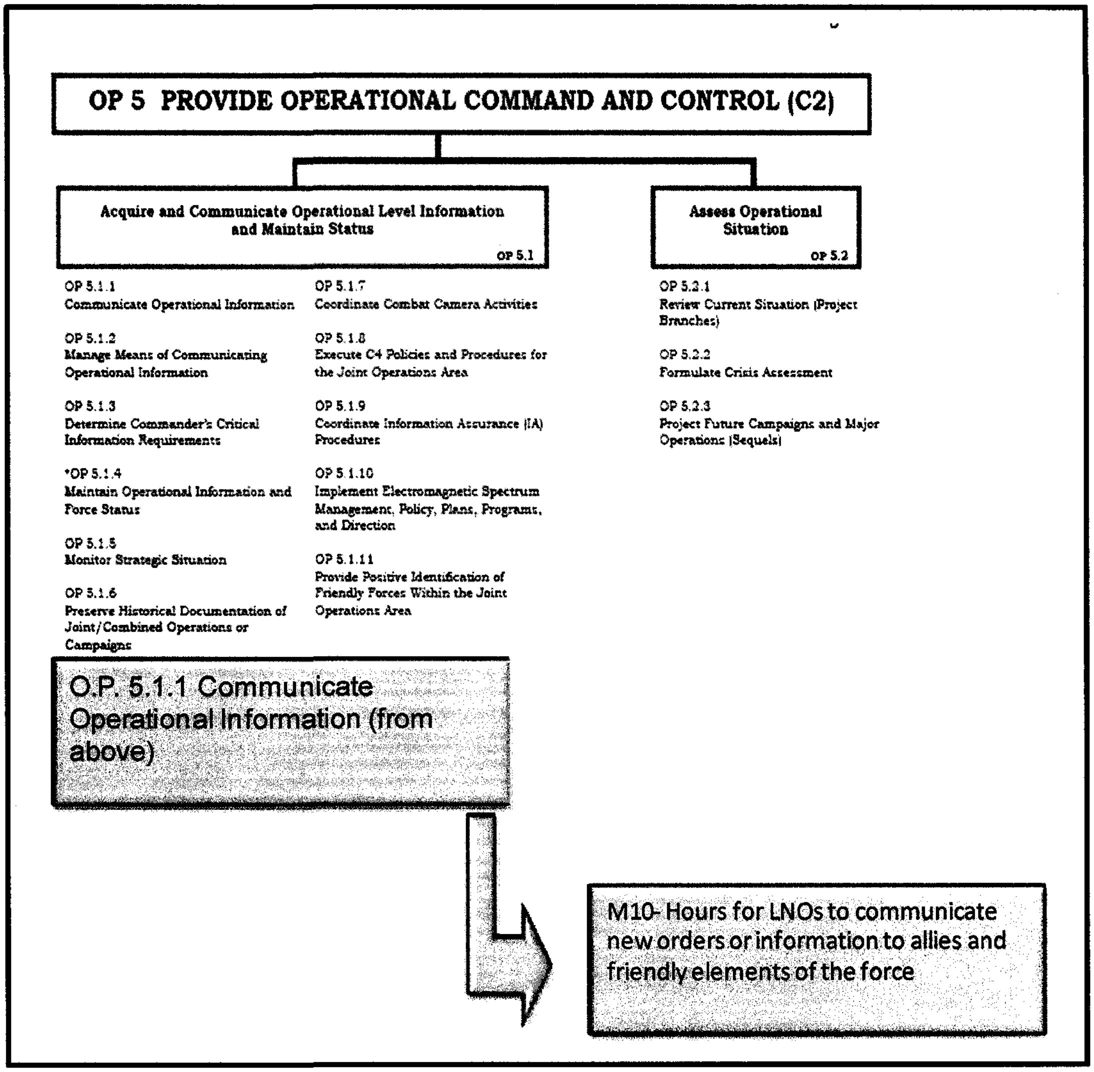

Figure 6: Joint Task for Operational Mission to One Relevant Performance Measure (1)

Thus, it would make sense to use a common language, if one existed, to describe the tasks required. Figure 6 illustrates how a task at the Operational level may be decomposed to a set of measures, not all of which apply to only that particular task but the measures used to set the standards for task accomplishment.

Since the implementation of the Goldwater-Nichols Act in 1986 (6), the ability of Service staffs to integrate and operate effectively with other Services, Allies, and agencies has improved. The UJTL is the document that provides the basis for understanding the tasks required at this level, and serves as the basis for not only completing mission sets, but also as the foundation for various Joint Capabilities. The most logical and potentially beneficial method to quantitatively assess the Operational Level staff is to use the framework provided by the UJTL.

The idea that a common language for training and readiness reporting would be beneficial is not new. Tillson advocated basing the DoD readiness reporting system on the UJTL in 1996 (24). Even with the extensive nature of this document, all situations, all possible operations, and all possible measures, standards, and conditions are not captured. The development of additional Joint Mission Essential Tasks (JMETs) or measures is one way to address this issue. The other is to supplement the UJTL with adjunct observations specifically relevant to key tasks, which is a method being utilized by the U.S. Air Force regarding its Operational Level staffs. Despite the points raised in the preceding paragraph, a focused look at the UJTL provides an extensive platform for use in evaluating Operational Staff performance (37), (38). At this writing, there are a minimum of four established efforts to utilize or develop operational templates for developing training scenarios, evaluating staff performance, and even reporting staff capabilities. The Commander Second Fleet (C2F), the U.S. Air Force Blue Flag[10] program, the Standing Joint Force Headquarters (SJFHQ), and portions of U.S. Pacific

---

[10] Blue Flag is a Operational Level exercise program used by the U.S. Air Force to prepare its operational staffs for deployment. This exercise relies heavily on synthetic environments.

Command are either actively developing or routinely using portions of the UJTL as the basis for Operational Level readiness reporting, training evaluation, staff training, or all three (39). A later portion of the dissertation will detail a method to utilize the information available in the UJTL to assess Operational Level staff performance based upon the UJTL.

2.1.4 Proposed Contribution to the Current Body of Knowledge for Performance Assessment

As noted in Section 2.1.3, there is a qualitative method available to evaluate the performance of operational level staffs. There is also an emerging requirement to report Operational Staff readiness in quantitative terms using the Defense Readiness Reporting System, DRRS, but there is currently no method, certainly not one common to all Services, COCOMs, or Operational Commanders to quantify readiness, regardless of the training environment available.

The operational template utilized to evaluate U.S. Navy Strike Groups at the tactical level, expanded by the use of methods proposed by McGinnis and Stone or Startin would appear to provide a means of quantifying performance. Neither of these methods can be directly applied, but with modification one or the other and perhaps both could be useful.

The McGinnis and Stone method was a proposal, but there is no evidence that it has been implemented or adopted for routine use. One goal of this research is to develop the families of tasks and subtasks, as well as the required weights, to perform assessments for the major processes for an Operational level staff.

As noted earlier, Startin's method anecdotally appeared to very appropriately model the learning curves for a staff in the warfare areas necessary for a U.S. Navy Strike Group to achieve certification. The processes necessary for an Operational level staff to succeed differ from warfare areas, and it is not realistic to apply the process, outcomes, and circumstances factors devised by Startin in this context. The skills necessary for a

staff to succeed have similarity to those utilized by Startin, and some of those will be used where applicable.

During Startin's research, the determination of weighting factors to apply to the LDFs and the normalizing factors was done empirically, and though this could be repeated, a more academically rigorous method of determination of the weights using Analytical Hierarchy Processing (40) was investigated and applied during this research. There will need to be two sets of weighting factors since staff performance can be determined in terms of either the organizational elements of a staff or the major functions, assessment, planning, and execution. The details of what constitute organizational elements or major functional areas will be addressed more fully in the Performance Assessment Methodology section of this dissertation.

There should be at least one method that links the organizational element performance results to the major functions. One of the issues that will be investigated during the course of this research is devising at least one method to portray such a linkage and demonstrating the validity of the linkage.

Though there are performance assessment methodologies and schemes in use in the military and other sectors, there seem to be two major categories of methods. There are the very subjective assessment methods which rely exclusively on the view of the assessor's knowledge, which have proven effective for very senior mentor relationships or individual training. There are also the typically binary methods which are quite suitable for initial training on an individual basis. As discussed earlier in this chapter and section, there is some limited collective learning assessment capability currently being used within the U.S. Navy to evaluate Strike Group performance as well as shipboard warfare capability. The ability to effectively assess an organization at the operational level, using a mixture of what are normally considered "soft" measures dealing with planning, and assessing or decision making with "hard" measures dealing in absolute factors such as time or percentage has not been attempted routinely or shown to produce results that reflect the ability of the staff. The development and demonstration of the feasibility of such an approach, which also permits the leadership to tailor the measures,

weight factors, and skill requirements to a specific situation in relation to a known set of standards, such as the UJTL, has not been accomplished successfully to date and is the goal of this research.

2.2 **Cost Assessment**

2.2.1 Previous Work on Costing and Cost Assessment

There are a great number of references that discuss cost accounting, cost management, and Activity Based Costing (ABC). An often-quoted author whether the discussion is costing, training, or return on investment is Kaplan. In Cost and Effect, he discusses the fact that the new environment in business since the mid-1970s demands more relevant cost and performance information on the organization's activities, processes, products, services and customers (17). They propose a four-stage model for cost and performance measurement, and observe that many companies can achieve what he terms as Stage II systems, which meet financial reporting requirements. They also note that many also attempt to go directly to what he terms as Stage IV systems which integrate cost management and financial reporting without doing the necessary work to implement a Stage III system which includes operational feedback systems that provide operators and all front-line employees with timely, accurate information, both financial and non-financial, on the efficiency, quality, and cycle times of business processes (17). A theme, which seems to resonate in any academic treatment of costs, performance, and ROI, though in an entirely commercial context, is that the ability to measure the performance of nonfinancial processes, especially involving services is a key to costing accurately. Kaplan devotes a significant portion of his work discussing the use of ABC in service industries. He presents a case study on government agencies, which notes that most opportunities for ABC are found in process improvement. He also noted that privatization was an area for investigation. The distributed synthetic training environment, and in many cases even the live environment (when one considers the use of Commercial Air Services (CAS) as a means to provide opposition air capability) are dependent upon private firms who can deliver products in support of training. The

discussions by Kaplan regarding privatization need to be reviewed in understanding costs for training to ascertain the effectiveness of the support contractor's ability to deliver services to the customers, in this case the Services and specifically the CJTF or FCC staff being evaluated during a training event

A methodology that formalizes cost accounting and the relationships of costs to activities is Activity Based Costing/Management (ABC/M). Gary Cokins is one of several researchers who have investigated the applicability of ABC/M to a variety of organizations, specifically applying ABC/M to the Federal government. He cites two US military examples of successful ABC/M implementation in his work. The US Army at both Ft. Riley and at Ft. McPherson embarked on ABC/M projects to solve complex problems. The issue at Ft. Riley was the need to define the true manpower requirements necessary to carry out functions within the Directorate of Logistics. The majority of the effort centered on reviewing processes and re-engineering those as well as adopting measures that might provide for continuous improvement. The U.S. Army Forces Command (FORSCOM) at Ft. McPherson needed to revise the procedures in place to provide more effective base support services to its eleven subordinate installations. Again, the effort was primarily focused on reviewing processes that involved needs such as child care, recreation centers, family housing, and fire protection, but the approach resulted in the adoption of new software, and re-engineering existing processes. One very interesting note was a decision to forego implementation of additional functional models when the ones already implemented utilized 90% of the allocated resources (15). While the efforts may not be similar in scope or function as will be required for determining ROI, the guidelines and approaches used in this study supply insight as to "how much is enough," as well as the need to adopt new techniques (in this case better software) if there is a clear advantage. Cokins also cites examples where the Canadian Army as well as seven other Federal or Municipal agencies implemented ABC/M (15).

Cokins' discussion notes that most senior leaders understand that any decision regarding cost affects the future. Further, he notes that the government and defense organizations must think of themselves as service providers and view cost management

accordingly. He provides a list of key players who are very much affected by ABC/M and a rationale for each key player to either be educated further on the benefits of ABC/M or the need to implement ABC/M since it provides a means to understand past costs, not just past spending:

- The civil service worker or military member might simply prefer the status quo or whatever may be a little bit better for him,
- The taxpayer prefers to be taxed less,
- The user of government services desires more and higher quality service,
- The functional manager is defending the existing level of his resources and fiscal budget (15).

One point discussed by Phillips was the intangible measures or "soft data" that are so difficult to capture. Cokins' view regarding intangible measures is very emphatic as quoted below:

> "The argument that any department doesn't have outputs is bogus. Managers and workers who think, plan, and provide direction conclude that since their work deals with intangibles, not things, then there is no definable output from their work. But the outputs can be intangible. Many are. What is the output of a university education? Is it the diploma? Is it each professor's course? Is it the learning by each student? These may all appear to be intangible. But the financial cost for each one is measurable (15)."

Cokins provides three examples where ABC/M is beneficial, which apply directly to use by a CJTF or FCC:

- Performance measurement – to provide some of the inputs to weighted and balanced scorecards designed to improve performance and accountability to taxpayers,
- Process improvement/operational efficiency – to optimize resource use and, at times, to serve as a key to an agency's survival,
- Aligning activities to the strategic plan – to correct for substantial disconnects between the work and the service levels than an

> organization is supplying and the activities required to meet the leadership's strategic goals (15).

Cokins' discussion of Activity Based Costing (ABC/M) in government settings takes a more aggressive stand on the issue of quantifying costs in service organizations. In fact, he asserts that without considering the costs associated with thinking, planning, decision making any estimation of cost is "bogus" as is the assertion that any of these activities are not valid for cost estimation (15).

Cokins' review of the use of ABC/M in the commercial sector and provides examples, both in method and in the form of case studies to illustrate applicability in many governmental settings. A very key portion of his work is the list of Federal Regulations, instructions, memos, and processes in effect (notably only the Government Performance Review Act (GPRA) from 1993 and Government Management Results Act (GMRA) from 1994 have been in effect for more than ten years).

Oliver's 2004 strategic assessment of cost systems discusses the fact that although cost systems initially served to aid in the manufacturing sector, they now perform functions that are much broader and in fact there is application to service organizations as well (41). She notes that cost systems serve four major purposes in an organization:

- Financial Reporting,
- Cost Measurement,
- Performance Measurement,
- Decision Support.

As with the other authors surveyed, Oliver's view on performance management is not merely about "how much is produced," or analyzing the past. She notes that a cost system should serve to make the critical link among operational measures, resource utilization, and costs in order to influence the future. It should be a tool to help manage organizational performance (41). Once again, the need to assess operational performance,

which is the "product" of a CJTF or FCC staff, is the key ingredient in building a case regarding costs and resources, effectiveness of the training system of choice, and eventually ROI. Oliver notes that traditionally cost systems are severely lacking as tools that help manage organizational performance. Her overarching principle for designing a strategic cost system revolves around understanding what the most important functions of the organization are and how much is willing to be invested to make it work. Throughout the book there are excellent examples of strategies, cost calculations (including material, labor, cost rollup) capacity utilization, labor utilization, and tools such as the use of Monte Carlo Simulation to set standards. While most of the examples revolve around the commercial manufacturing and service sectors, the issues involved in using distributed simulation have many common attributes and these strategies, methods, and tools will be reviewed for applicability throughout this research effort.

Wright and Belcourt's 1995 article highlights their belief that determining the costs and benefits of training is not well understood. They propose a practical approach in addressing the processes and steps necessary to develop a decision-making system (42). The principles espoused by the authors include the need to treat training costs as an investment like any other and then assess the performance of training using all costs available. The costs, both direct and indirect, defined by the authors then provide a basis for looking at alternative providers. One aspect of the costing discussion presented by these authors is the presentation of a costing methodology in conjunction with a scenario that clearly illustrates the possible benefits and tradeoffs for potential training options. The authors discuss qualitative assessments as well as how credibility for entering assumptions and information plays a crucial role in the overall decision-making process. Finally, the authors give examples for reducing costs, corporate approaches to absorb training costs, and the need for strict record keeping. The issues of credibility, qualitative assessment, cost reduction, corporate cost sharing, and record keeping are noted as being extremely culture sensitive; thus, they must be continually addressed if the processes and methods described are to have any chance of success.

Three other noteworthy points were raised in the Wright article. The first dealt with the concept of how to insure credibility for cost/benefit analysis. The example cited a case where corporate executives assigned a value for turnover costs without consulting any other representatives in the company. Though a cost/benefit decision was reached, it would not be unusual to find that in some later analysis, the benefit would not be as great as forecasted because the actual turnover cost was higher than assumed. The risk of acceptance (or, more aptly, providing a compelling argument to management for acceptance) of a specific cost by management from an entity outside that limited group is a cultural issue that is addressed very neatly by both Cokins and Schein. If a manager is defending his allocated resources, a "back room" estimate of any sort will not suffice (15). Just as important, as discussed by Schein, if management does not accept or see a need for cultural evolution, it will occur but in what might be a less desirable form. Additionally, trainers must ensure that the organization sees how training is important to the culture or an organization and how the organization's culture can deal with any challenges that are a struggle (43). Both these views address the concerns raised in the Wright article and address it in a much larger context such as the Federal government or in an industrial environment.

The authors provide some possible methods to reduce costs for training. As has been stated in previous sections of this proposal, unless a specified cost is attached to many of these items, it would seem that the savings would be minimal in most cases. Travel (shown as "off-site"), outside consultants, or "prestige" trainers are very often the most expensive costs encountered in training and of the items shown on the list, the ones that really have the potential to save any organization money (42). The other costs, understanding that in any specific instance there could be exceptions, would probably not be constraining except to a very small organization.

The last issue in the article which seems to be contrary to the entire philosophy of the article concerns the statement that," Without measurable benefits, the entire human resources development function loses credibility (42)." Training, the costs incurred, and the efficiency of any organization is not the sole responsibility of the Human Resources

Department (HRD). As noted by a number of other authors cited at various times in the preceding discussion, training, costing that activity, determining ROI, and then having a culture that accepts all of these issues is an organization wide issue. More specifically, leadership must be involved; not only in assessing value, but also in administration and ensuring the organization's philosophy supports training. Fortunately, Wright and Belcourt discuss management's role and the need to overcome either institutional or cultural bias regarding the costs and benefits of training, so perhaps the allusion to HRD as the only entity to lose credibility was an oversight in the article.

Kraiger discusses the costs of training and though the areas he stresses as those requiring attention would seem obvious, they are often overlooked. Examples of the types of costs he recommends tracking are:

- Direct costs include training materials, instructor payments, use of space and equipment, and travel expenses for participants,
- Indirect costs are associated with maintenance of the training function, overhead, and postage, shipping, and so on,
- Development costs may include payments to external consultant s or the costs associated with the project linked back to the training development staff.
- Finally, costs associated with trainee attendance are calculated, with the assumption being that if trainees are being paid to attend training, then they are being paid not to contribute directly to the production of goods and services by the organization (44).

Kraiger's discussions regarding costs also led directly to the development of ROI, which will be presented in that discussion later in this proposal.

Other than "classical" accounting or ABC/M, are there other methods that might be useful in applying costs to activities, especially in the context of training exercises as they relate to a METL framework? The answer to that question is yes, and some work done by Martha Nelson and Dennis Bely within a Missions and Means Framework may have potential for solving both the costing and cost relationships to METL questions.

In 2005, Nelson and Bely published a report regarding Live-Fire Test and Evaluation (LFT&E) Strategies (45). The motivation for their work arose from the Army's need to change previous processes for conducting live fire tests which required updating from Cold War practices wherein LFT &E was used to determine survivability and lethality of an autonomous combat platform to that of a System of Systems (SoS) context in an integrated and information-centric battlefield. Two key issues bearing investigation that are relevant to this proposal were:

- How should LFT&E planning and execution change in light of the tactical SoS tactical doctrines?
- How should LFT&E be designed and conducted to evaluate LFT results so decision-makers can ascertain mission success in the joint environment (45)?

Nelson and Bely provide a supporting case for the change in approach that parallels some of the discussion presented thus far regarding performance assessment. Traditional platform-centric testing strategies focused upon the extent to which a platform retained its battlefield combat utility. Little consideration was given to the critical issues surrounding the complimentary capabilities of other platforms that are part of the SoS or to the platform's role in the completion of tasks linked to specific missions in the joint environment (45). This is analogous to the isolated nature of training currently conducted by the Services, even though the UJTL provides a joint framework that could be used, as proposed by Tillson for the Services to relate tactical _METLs to the Operational level. The reporting of Service training and readiness without a joint environmental context is useful but not nearly as complete as possible, since the mission or capability that is achievable by the unit or staff being reported upon is missing.

2.2.2 State of the Art in Costing Methodologies

Although there is a common framework for depicting the operational capabilities of a staff, the ability to capture the costs required to train such a staff is comparatively elusive. Beginning in 1993, with the Government Performance Results Act (GPRA), and followed by the Information Technology Management Reform Act (ITMRA), a Undersecretary of Defense memo, the first Quadrennial Defense Review (QDR) in 1997,

and finally the National Defense Panel (NDP) of 1997, the requirement to not only monitor costs, but to implement cost schemes that emphasized all areas of military performance has been clearly stated (15). The USD memo specifically directs the DoD wide implementation of ABC/M, noting that unless it is most effective when pursued on a wide basis and not just on narrowly focused issues such as weapons systems life cycle costs. The GPRA, ITMRA, QDR, and NDP all point to the need for DoD wide cooperation in acquisition or at least interoperability of systems, especially IT systems. Thus setting up a methodology to account for the costs of systems, and even the costs of training events would seem to be something that is at least defined at some level in each Service. The ability to explicitly extract costs or show similarities in costs may not be clearly delineated, but as long as the intent of the regulations has been implemented, the ability to build some sort of algorithm for costing should be achievable.

Inquiries made to the Senior staffs of the U.S. Air Force, U.S. Army, and the U.S. Navy's Commander Fleet Forces Command illustrate that while costs for training are being tracked, there is no uniform method. Any method currently used is not directly related to the UJTL, which makes the depiction of costs in any training environment, including at the tactical level difficult. Efforts to track costs, especially in light of the rising costs of fuel, transportation, and infrastructure are increasing. These methods rely on spreadsheet methods within the individual Services, which once again precludes easy translation between Services or a uniform method of portraying the specific operational and Joint Capabilities that have been achieved for a specific expenditure of funds.

2.2.3 Costing Methodology Issues Requiring Resolution

Fazal completed one of the few academic efforts to capture costs as it related to both military training and computer-based instruction 1996. He reviewed the implications of using Computer Based Instruction for both higher education and military training. He examined the effectiveness of computer-based instruction (CBI) compared with conventional instruction using a meta-analytic research synthesis of evaluation studies conducted since 1990. Fazal noted that from his analysis, effect sizes were greater for higher education than for military training (46). Unfortunately, the type of training

examined by Fazal, and the measures used would be insufficient even for the types of complex simulations that individuals are currently subjected to.

There was no evidence uncovered by this researcher during the preparation of this dissertation, which attempted to measure the effectiveness of training and directly relate the effectiveness of that training to costs, though there have been limited works that discussed staff processes or the development of simulation systems. Suantak's work in 2004 discusses simulations at various levels including the Corps Commander level, which is applicable for FCC purposes (1). Unfortunately, her work discusses formulation of the simulation and not the ability to assess it or to link costs to performance, which is not particularly useful for this investigation.

There are several authors who have introduced useful frameworks for cost accounting. The acceptance of a framework which captures the detail already delineated by the authors to date, especially Nelson, Oliver, Cokins, and Wright, and then extends these efforts to the UJTL would be a major step providing all the Services at a minimum, a common structure for portraying the improvements in readiness (as determined from performance assessment) realized from a given expenditure of funds.

Cokins' listing of key players affected by ABC/M discussed in Section 2.2.1 provides the motivation to clearly understand costs incurred during military training. The warfighters cannot just accept the status quo in the very dynamic combat and geopolitical environment for which they must prepare. Though it would be foolish to disregard lessons learned from past contingency operations, basing all training on "the last war" has historically proven to be costly. Solutions which lead to improving the quality of training in terms of realism, more rapid and objective feedback, and that are more time efficient are required in this new era of short noticed deployments. Taxpayers understand the need to adequately fund readiness, but also desire the lowest tax bill achievable. Finally, the Service chiefs and Combatant Commanders must have a credible method of defending the resources needed to ensure that US forces deploy with maximum capability.

Though there are some clearly quantifiable activities that occur at a CJTF or FCC staff, the vast majority could be described as thinking, planning, and providing directions. In fact, if those three activities are not pursued, what is the point of having levels of command above those executing missions at a tactical level at all? Whether those activity "costs" are initially captured as dollar figures or metrics which then translate to time and thus costs which can be captured as man-hours, network usage, or in some other form is unimportant, the fact is that a method to develop a set of metrics to capture these costs is key to all three endeavors, costing, assessment, and ROI.

The methodology presented in Nelson and Bely's report provides additional insight into assuring that activities commonly found in a military context are documented. The complementary nature of this work to the work completed by Wright and Oliver provides a method for assuring that the activities typically encountered in military training can be accounted for while retaining both the language in the Wright article and the strategic viewpoints expressed by Oliver. The methodologies and insights provided by all the authors noted in this section will be used in various combinations to construct a more formal and repeatable costing methodology for operational level training exercises. There is little doubt that costs are being attributed to these events, and that in almost every case, very detailed specifics concerning various aspects are being recorded and then used to show training effectiveness, but using the concepts expressed by the authors cited throughout should permit a more uniform structure as well as permitting costs to be related to training accomplishment measures vice just exercises overall.

2.2.4 Proposed Contribution to the Current Body of Knowledge Linking Cost to Training Accomplishment

The costs of training, at least in terms of cost per major event have been quantified for live, virtual, and constructive training. The costs for some specifics such as fuel costs for individual platforms, costs for contractor support, costs for target range use, and even costs required to determine environmental impacts from the use of ordinance have been quantified. Even with these sorts of details, the costs necessary to

achieve essential training objectives, either in terms of individual warfighting area proficiency or at the Operational Level in terms of a staff's major functions or organizational elements, is still not well defined.

The JMETL framework used to evaluate performance must also be used to link the costs associated with achieving mission sufficiency. The UJTL is, for the moment, the only common language available for such purposes. Approaching costing using the UJTL provides several advantages, regardless of training environment that are presently not available.

The best case for linking costs to training would be to link costs to the satisfactory completion of individual measures subordinate to the JMETs. Thus the cost to accomplish any specific task could be directly tied to resources. As depicted in Section 1.1.1, it might even be possible to show a relationship to the achievement of a Joint Capability as well.

One detriment to assigning costs to specific measures is that this assignment process may not be practical. Even if a linkage for many specific measures could be established, how would these costs be compared to costs which could only be realistically assigned to a single JMET category? Additionally, it may not be practical or possible to establish such a linkage in all cases, especially for those sorts of costs that were termed intangible by Cokins and Phillips.

The skills used to "bin" measures address traits, which are generally considered intangible such as authority, leadership, competency, or initiative, may provide a better opportunity to assign costs. This could be the case since there are a number of measures that must be combined to create the bin, and this combination may prove more amenable to assigning costs. An example is in the area of intelligence. The competency of an intelligence support team is determined, at least in part, according to its ability to properly assess damage effects. There are very specific costs, especially in a synthetic environment, required to create the imagery, provide analytical support, and finally to communicate information. The overall effect of these costs provides an opportunity to

assess the competency of the intelligence team. It also provides an opportunity to assess leadership and initiative. There are systems, support personnel, and scenario injects regardless of the type of training being assessed that permit evaluation of areas that have generally been considered intangible. The entire focus of the Operational level staff's efforts, especially in terms of major functions, is largely what has been traditionally termed intangible, so devising a method to assign costs to all aspects of staff performance is important. Assigning costs to skill bins is more general than the assignment to specific measures, but introduces another potential discontinuity since there might well be cases, as described above, where other costs can still only be realistically assigned against a JMET category.

The ability to assign costs, especially in terms of direct costs, personnel costs, and even one time or event specific costs is not nearly as difficult; though it appears, at least at this writing, that these costs will very likely be assigned to the JMET level vice measures or skills. The assignment to JMETs is acceptable since there will be targeted JMETs for any training exercise regardless of the LVC nature of the training scenario. The ability to clearly define those costs associated with a live scenario as opposed to those required in a synthetic one and the savings achieved, if any, is a focus of this research and an important factor in determining ROI.

It is possible that costs might only be assigned to events, which has utility, but this would greatly diminish the utility of the analysis since it would be either difficult or anecdotal to attempt to compare costs to other events unless they were nearly identical.

None of the foregoing discussion is particularly revolutionary, even assigning costs to the "intangible" aspects of staff activities has been discussed by Cokins and Phillips (15), (16). The unique nature of the costing methodology proposed is that costs have not been reliably assigned to JMETs at the Operational Level, there has not been a definitive effort to assign costs to the "intangible" aspects of staff performance, and there is no uniform method to compare costs in various training environments, notably live and synthetic, to each other to really understand the advantage or disadvantage of a given environment in achieving operational objectives. There is an intuitive sense, and there

has been some anecdotal calculation of the savings achievable in a synthetic environment, but specifically attributing costs to specific training objectives as a basis for such a comparison is not yet in place.

Regardless of the ultimate method chosen to assign costs within a UJTL framework, it appears that some form of aggregation, either in cost categories, which would still be very general, in skill bins which may be too specific, or by assigning costs to specific JMETs is the solution which will prove most effective for determining the costs required to complete training at the Operational Level.

Accounting for resources expended for specific uses is not a new concept, and even being able to portray the percentage of any organization's budget which was allocated for training is done in many sectors. What is currently not well done is specifically and formally accounting for the costs associated with specific tasks which are required to attain a required level of proficiency to permit operational employment. The Measures and Means Framework efforts of Nelson and Bely (45) established a feasible method for such an approach, but the approach has not yet been extended to a larger context. By developing a very formal method of categorizing costs and linking these costs to specific METs, the goal of this research is to provide a first step in quantifying what operational readiness was achieved for what was allocated.

The next section of this dissertation will discuss the implementation of ROI in various sectors. Though ROI implementation in the public sector is typically slower than in others, the implementation of either ABC/M or some other system as rigorous in the military is a very necessary step in developing meaningful ROI.

2.3 **Return-on-Investment (ROI)**

2.3.1 Previous Work on Return-on-Investment (ROI)

A leader in training evaluation and the development of measures to capture performance improvement is Donald Kirkpatrick. He has been writing on the subject since the 1950s and is probably one of the preeminent authorities on training evaluation.

In 2006, he revised a seminal work on the subject of training evaluation. This revision addresses ROI. Again, the areas that he highlights as key to evaluating an organization are those that are commonly attributed to CJTF or FCC staffs and also capture the essence of the costs of interest:

- What results are we trying to accomplish? These results can be stated in such terms as production, quality, turnover, absenteeism, morale, sales, profits, and return on investment (ROI).
- What behaviors are needed to accomplish these desired results?
- What knowledge, skills, and attitudes are necessary to achieve the desired behaviors (12)?

Kirkpatrick also discusses why it is important to evaluate training. None of his observations are startling, but all are applicable to US Forces and especially to the development of a methodology for ROI:

- To justify the existence and budget of the training department by showing how it contributes to the organization's objectives and goals;
- To decide whether to continue or discontinue training programs;
- To gain information on how to improve future training programs (12).

Although there are four levels to Kirkpatrick's evaluation model, we are principally interested in what he terms evaluating results. Of all the levels discussed, this is the most difficult. As discussed by Phillips and Cokins, Kirkpatrick notes that evaluating results is primarily the evaluation of organizations, which require measures that are largely subjective and often ill defined. Some approaches to developing these sorts of measures have been discussed earlier, but Kirkpatrick's comments only re-enforce the need to develop more specific measures for the FCC and CJTF staffs.

Some of Kirkpatrick's guidelines, like those of other authors cited thus far are applicable to almost any organization, but may be far more difficult to implement in a CJTF or FCC context due to the very limited nature of training exercises as compared to

processes which are encountered in an industrial or educational setting. He also discusses the need to allot time for results to be achieved which is very much akin to the requirements to implement an adaptation to the W. Edward Deming's model for Total Quality Management, which was termed Total Quality Leadership (TQL) (47) which was introduced to the US military in the mid 1990s. Perhaps, with reconstruction of previous data, if a reasonable formulation can be achieved, some time for performance can be accommodated to ensure achievement of useful data for ROI purposes. Selected guidelines from the discussion include:

- Measure both before and after program if practical,
- Repeat the measurement at appropriate times,
- Consider the cost versus benefits,
- Be satisfied with evidence if proof is not possible (12).

Kirkpatrick discusses e-learning (in the same context as defined earlier by Sloman) and methods to evaluate this newly emerging tool as well. One very promising use of e-learning would be in pre-exercise evaluation and then in remediation in the case of specific performance deficiencies (10). The Army and Air Force Exchange Service (AAFES) is cited in a case study by Kirkpatrick as a specific example of a military activity that used the four levels of evaluation in training to improve a process (12). Though the issues are more similar to a commercial marketing situation, the key is that a military organization was able to adapt his approach to its situation. It is also true that the vast majority of personnel involved was civilians, but the leadership, which is often the most difficult group to persuade regarding the implementation of new or at least different methods, did ascribe to the process, so the model is valid.

There has been extensive work on ROI as it applies to non-military applications. In his 2003 book, Return on Investment, Phillips develops a case for and methods to compute ROI. He begins his examination with a statement that applies to any organization, "While the payoffs are assumed to exist and training appears to be needed,

more evidence is needed, or training funds may not be allocated in the future (16)." This statement captures the very nature of the training issues faced by US forces.

Training for the entire scope of possible contingencies that are liable to occur is a given, how to accomplish that and what cost is too much and thus prohibitive is not easily determined. Two further examples of the potential for ROI are provided for consideration. Phillips depicts the progression of ROI within various sectors as reconstructed in Figure 7. He notes that use within the Public sector has increased in recent years, but actual use as a tool for training and budgeting within the military is still limited (16).

Movement within the Sectors

Manufacturing Sector
↓
Service Sector
↓
Health Care Sector
↓
Nonprofit Sector
↓
Public Sector
↓
Educational Sector

Figure 7: Progression of ROI Implementation
Reconstructed from Phillips, page 4

A different view of the applicability and desirability of ROI is shown in Table 1 (16). Again, though not specifically cited, the military certainly falls within the organizations, which rate each metric as important and certainly wants to see the impact of training experiences on performance.

Metric	Percentage of organizations rating metric as important or very important (wish list)	Percentage of organizations indicating utilization of metric (doing list)
ROI in Training	78%	11%
ROI in Development	62%	2%
Impact of Training Experiences on Performance	88%	27%

Table 1: ROI Opportunities

Reconstructed from Phillips, page 6

A key part of discussions by several authors, including Kaplan, Kirkpatrick, and Cokins is the fact that measurement without evaluation of performance, whether that is in the form of a balanced score card or by level, is meaningless (17), (12), (15). Phillips adds to this discussion as well; notably, he states:

> "Evaluation is an integral part of the design, developments, delivery and implementation of programs; the implementation of a comprehensive measurement and evaluation process usually leads to increased emphasis on the initial needs analysis (pre-testing/ team evaluation); organizations with comprehensive measurement and evaluation have enhanced their program budgets; a comprehensive measurement and evaluation process, including ROI, can be implemented for about 4 or 5% of the direct program budget; technology is significantly enhancing the measurement and evaluation process, enabling large amounts of data to be collected, processed, analyzed, and integrated across programs (15)."

The other interesting view expressed by Phillips is the addition of another level of evaluation, which is ROI beyond the model discussed by Kirkpatrick. Phillips does warn that there are some common errors that either defeat or considerably weaken efforts to

implement ROI. To avoid those pitfalls he proposes several criteria for an effective ROI process. The criteria that most directly apply to any effort, which the military embarks upon, are cited below:

- The ROI process must be simple, void of complex formulas, lengthy equations, and complicated methodologies. Most attempts fail by not meeting this requirement,
- The ROI process must be economical and must be implemented easily. The process should become a routine part of training and development without requiring significant additional resources,
- The assumptions, methodology, and techniques must be credible,
- The ROI process must include the costs of the program. Omitting or underestimating costs will only destroy the credibility of the ROI values,
- The actual calculation must use acceptable ROI formula. This is often the benefits/cost ratio (BCR) or the ROI calculation, expressed as a percent (16).

All of these are achievable and applying these criteria to any training exercise effort would be substantial progress towards quantifying what the CJTF or FCC staff is gaining, whether the exercise was live or synthetic.

The data flow represented in Figure 8 is a partial representation of Kraiger's Figure 11.2 (44). This depiction re-enforces the points made by Kirkpatrick regarding the value of assessment and even yields potential insight for ROI. Kraiger also portrays additional indicators of organizational payoffs in Exhibit 11.3 in the same chapter that also gives description, concrete examples and references for each focus of evaluation. He depicts the indicators in three dimensions.

He portrays the layers to be used by learners, decision makers, and marketing as methods, focus, and organizational payoffs. The analogies for the military context and this proposal are ROI, performance assessment, and combat readiness.

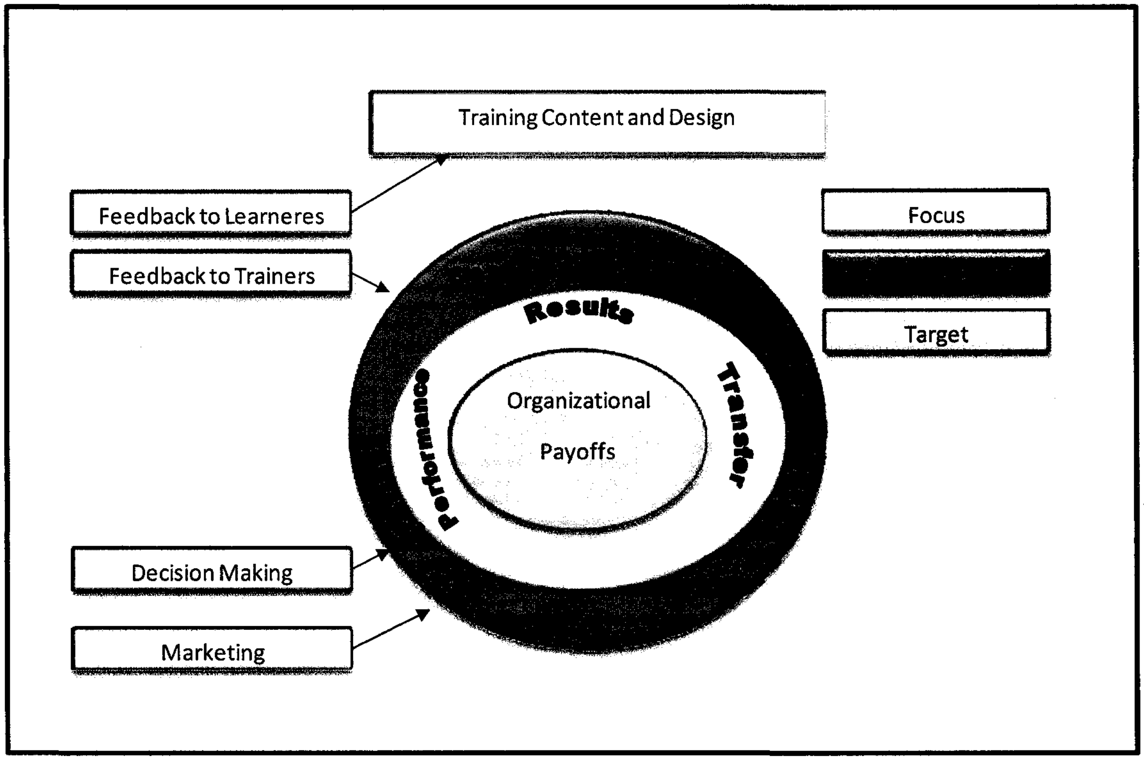

Figure 8. Targets of Evaluation (44)
Reconstructed from Kraiger, Page 343

Equally useful for application to a CJTF or FCC are the Results and Financial Impact dimensions (44). Kraiger's discussion of results compliments the work done by Kirkpatrick and also offers additional viewpoints for consideration and implementation including approaches to ROI as well as the possibility of modeling to estimate the impact of training, which readily translates into using assessment as a means of measuring training effectiveness. Even though the preceding discussions would appear to provide a clear and concise framework for determining ROI, there is still substantial there is still substantial confusion as to what constitutes ROI as discussed by both Dust and Worthen (48), (49). An excellent example of a generalized scheme and rationale for ROI in industry is provided by Worthen in a 2001 article in CIO magazine (49).

One interesting note in this article is the fact that Worthen notes that the monetary savings may not be as great as expected (though for a comparison of live vs. synthetic military training, fuel alone should provide substantial monetary benefits), and ROI may not be characterized by monetary savings alone. The discussion in this article is principally centered on IT investments and the need to better estimate the advantages those investments bring to industry.

Dust noted in his article "The Myth of Training ROI" that even with the completion of lengthy studies of formulas, the implementation of the appropriate structures to capture data, and consultation with subject matter experts, his expectations that determining ROI would be simple were not met. He further notes that even knowing that the key to gaining financial backing is proving that training will improve processes dramatically, after what appeared to be a smashing success, management doesn't see the value in increasing training resources (48). As viewed when the article was written in January 2004, Dust asserts that the biggest value training ROI produced was the sale of books and little more. He attributes this to several factors, attitudes about training and its value, the type of organization observed, and finally on training metrics. Dust views training effectiveness as the key, and that effectiveness can only be realistically judged by knowing pre-testing (training) levels of knowledge (LOK) and post-testing LOK, and comparing the difference (48). This is in concert with the views expressed by Rickover and Spurgin, and captures the compelling argument for assessing performance in the military context. Once a noticeable improvement in LOK is observed, then reducing the costs required to deliver training will finally lead to a formulation of ROI that is tangible. Not a traditional view, but one that has great applicability to a military, and especially an operational staff context and very suitable to implementation in a synthetic environment. Dust also differentiates between effectiveness, which he terms as the quality of training and the efficiency or quantity of training (48). Again, his measures are by increasing the numbers of personnel receiving the training for a given cost. This view again has immediate applicability to both operational level staffs and the use of a large-scale distributed synthetic environment as a measure of ROI.

Marcia Conner provides some more useful parameters for consideration in her discussion of ROI (50). Not only does the author provide a very practical approach to the calculation of ROI, she presents a listing of sources discussing ROI. Of particular note in her discussion is the introduction of factors that should be incorporated into ROI such as reduced travel time, reach across time and space, rate of performance change, timeliness etc. These are all factors that bear investigation for military training, regardless of whether it is live or synthetic.

There are several other examples of either discussion regarding the need for more focused ROI in terms of modeling and simulation or concrete uses of simulation that provided monetary evidence of the ability for simulation to provide ROI benefit. One specific example of a discussion which illustrates the need for greater focus on ROI when discussing Modeling and Simulation is presented in a paper entitled, "Increasing Return on Investment from Simulation" which was presented at the 2003 Winter Simulation Conference. The four panel members presented views on three aspects of ROI and simulation. Selected comments and viewpoints from this discussion provide insights into what had been done until that time, and many of these views still have validity for the research conducted for this dissertation.

Several references from the cited discussion provide specific examples of the use of Modeling and Simulation in a military context to provide insight into cost savings or ROI. Carter, Olden, and Gordon (51), (52), (53) have all written articles discussing how ROI can be achieved in a military context. Two of these, the Carter and Olden articles, are very specialized cases dealing with design and acquisition, but reveal little that is either revolutionary or applicable to a wider application such as training. The Gordon paper does address the use of simulation to significantly reduce costs during design, development, and testing for new systems. He also cites savings achieved in large-scale training exercises and depicts the cost savings or cost avoidance that could be forecast for similar training if M&S[11] was incorporated. The method he used to support his

[11] M&S acronym commonly used to refer to Modeling and Simulation

hypothesis was based upon the increasing reliance on M&S utilized in conducting the REFORGER series exercises since the early 1980s. In 1988, the large number of live forces, the tracked vehicles and tanks used as training aids, and the maneuver damage [12] resulted in an overall cost of \$73.9M. In 1992, the use of more robust M&S permitted the training of the same number of maneuver headquarters, but with 77,000 less personnel, nearly 8,000 less tracked vehicles and tanks, and no maneuver damage for \$19.5M. He also notes that the cost savings attributed to M&S amounts to at least \$9.0M per year for the U.S. Air Force as opposed to an approximate \$4.0M cost for M&S maintenance. Gordon also addresses the training for a battle staff by noting that in a typical Blue Flag exercise, if half of the sorties flown in support of the exercise were simulated, then a cost savings of approximately \$7.3M per day could be achieved (53). This effort only addressed a single service and though probably one of the first published papers regarding M&S capabilities, it is unclear whether or not additional benefit was achieved as a result of his efforts. It is also not clear what factors were used by Gordon to arrive at the results above, though it would appear that fuel costs and M&S costs were available. The results for training appear to be event specific, which as discussed in Section 2.2.3 is useful, but does not fully address the current needs of the Armed Forces. Gordon also notes that the advantages of M&S must go beyond costs, which was noted by Medeiros in his subsequent answer. This article does provide some support for the assertion that ROI is being used more effectively in the military than in other sectors (53).

A similar effort, albeit smaller in scope, was utilized in the mid 1980s to reduce the costs of ship design by using simulation in place of far more expensive tow tank testing for the Arleigh Burke destroyer class (54). The use of ROI as a metric might be better described in terms of culture as presented by Schein, which will be more fully addressed in a later section of this proposal. The essence of Schein's comments is that if

---

[12] Training aids include target ranges and the ammunition used. Maneuver damage is that damage resulting to the property and environment caused by the movement of large numbers of personnel and especially heavy vehicles such as tanks.

leadership does not evolve, then different sectors of the organization will (43). Clearly there are many opportunities for the military to use ROI as a simulation metric, but doing so as a matter of routine is just gaining acceptance by senior leaders.

McGibbon in the 1999 Revised State of the Art (SOAR) noted that,

> "The purpose of this Revised State of the Art Report (SOAR) is to provide more insight into the details necessary to demonstrate from a business, profit and loss, and senior management perspective the benefits of improved software management using software process improvement techniques. Software process improvement has received much attention in recent years. However, it has been very difficult to translate benefits achieved in one organization to another organization. The intent of this SOAR is to generalize and model the cost benefits one can achieve from software process improvement (55)."

The SOAR looks at possible alternative metrics such as reduction in time and effort to develop software, reductions in the defects in the software, and improvements in development team productivity. These concepts can logically extend to the use of large-scale distributed synthetic training, especially in the areas of scenario development, advanced concept testing, and the ability to reduce travel costs for geographically widespread staffs. The combination of metrics such as those proposed by McGibbon with performance assessment could result in measures for ROI which would address some of the areas the panel discussion members felt were lacking.

McGibbon is not alone in proposing that ROI may need to be expressed in a different method. Cross states that," Where you stand on ROI depends on where you sit. Different levels of management make different sorts of decision, so it's appropriate that they use different measures of ROI (56)." Later in the same article he states, "Rational decision makers look beyond an ROI that reduces everything to the lowest common denominator (55)."

2.3.2 State of the Art Regarding Return-on-Investment (ROI)

The present capability to portray ROI is limited at best. Efforts by Commander Fleet Forces Command (CFFC) have been able to show macro savings based primarily on fuel usage for ships involved in sequential exercises that are both live and synthetic. Such savings are dramatic but do not promise further development unless the personnel costs, travel costs, and other direct and indirect costs and NMET coverage can be cited.

2.3.3 Issues Requiring Resolution to Formulate Return-on-Investment (ROI)

An issue affecting the ability to measure ROI, especially in the area of training, and for this proposal, specifically operational level training, is the fact that as Dust and Worthen have noted, training doesn't fit into the traditional construct of ROI. Perhaps some redefining of ROI, or at least possible alternatives or acceptable formulations of ROI must be undertaken to properly characterize and measure ROI. Two examples of the need to change the definition of ROI both within the Federal Government and in industry have been proposed by McGibbon and Cross. Both express the view that ROI isn't just about investment and return as previously defined. The views are succinctly and stridently expressed, which should at least provide an invitation, if not a strong endorsement for leaders and managers throughout DoD to examine varying formulations of ROI if they are to make wise military and business decisions. Some examples of either some potential ROI formulations that are not presently readily available, but that could apply to the operational level as well as other sectors are supplied below:

- The ability to map cost to measures and the desired result (task completion or operational capability) to predict the cost and time required to achieve the required training,
- The ability to make a value decision (having watched training improvement and knowing that human learning is logarithmic) regarding the sufficiency of training completed,
- The ability to make a decision about the relative benefit of additional training for an entire staff or just specific functions,
- The ability to choose specific scenarios, which apply to real world staff employment and match, costs to capabilities or tasks,

- The ability (in either a synthetic or live environment) to direct a JMESL[13] and then project costs or projected benefits,

- The ability to tailor training to a specific staff's capabilities or tasks.

As illustrated in the preceding discussions, whether from an academic standpoint or in a commercial/industrial context, templates or at least well-developed methodologies available to calculate costs translate what might be considered as intangible measures of performance into cost data. It is also possible to determine ROI for a wide range of activities, including services, which is the area of focus for any sort of military training, especially training which occurs in a synthetic environment. What is not readily available, at least at the unclassified level or in the public domain, is evidence or comparable methodology as implemented or commonly used within the Department of Defense. Specifically, no comprehensive algorithms or strategies have been found that specifically address a methodology that permits a repeatable and/or reliable formulation for ROI.

It would be unfair and unrealistic to assert that the Services are not meeting the mandates of the Federal Regulations stated earlier. In fact, it is entirely likely that each Service, each of the leading headquarters such as Air Combat Command or Fleet Forces Command has some method for costs, may have some formulation for assessment, and even some way of representing ROI. It is unclear whether there has been any discussion between any of the entities named which might permit using a common framework to capture ROI in a more uniform fashion to support each entity or the Armed Services as a group

As discussed in Section 2.1.2, the U.S. Navy has a methodology in place to evaluate the performance of its deploying Strike Groups, and their staffs. The basis for this evaluation system is _METL based, which could at least form a basis for

---

[13] JMESL – Joint Master Event Scenario List, the primary tracking method during exercises to coordinate time, events, and observers.

constructing the methodology to be followed since the _METLs used at the tactical level have very little application at the Operational Level. At the risk of being repetitious, perhaps in the extreme, there are sources for formulating costing, assessment of performance, and then using these results to construct appropriate representations for ROI. The methods, whether adapted from Oliver, Kirkpatrick, Phillips, Cokins or Kaplan will require modification, but all have relevance. The points raised by the authors reviewing the increasing use of synthetic training, the advantages it provides, and the need to expand its scope all either state or insinuate the need to provide relevant measures of effectiveness to ensure maintaining the capability.

In order to develop a methodology that might be applicable within DoD, or at least at the Combatant Commander level, several objectives must be met:

- The development of an effective model of costs for synthetic, live, and where applicable virtual training,

- The development of a set of metrics for staff assessment that capture the warfighting skills required to successfully manage the range of scenarios discussed that is also amenable to tailoring, weighting, and usable regardless of service,

- To use the results of both 1) and 2) to define a realistic and reliable tool to define ROI for any of the scenarios above which will permit defense of training scenarios and spending during budget reviews and annual exercise planning,

- To furnish commands such as JFCOM a possible template for exercise development which can be used to either initially certify or re-certify FCC or JTF staffs, with the ability to predict the costs, initial staff capability, and required skill sets needed to achieve certification.

There are some challenges that will have to be overcome to achieve the objectives identified above. Phillips makes the following observations, which must be accounted

for and overcome to succeed. It should be noted that although his focus was on sectors other than the military, these observations are very common to any military organization:

- Costs and time will be incurred when implementing the ROI process. These costs should not and are not typically excessive (3-5%),

- Lack of skills and Orientation by evaluation staff is common and most do not innately have either the basic skills necessary to apply the process within their normal scope of responsibilities. Measurement and evaluation are not usually part of the requirements for personnel assigned to an organization,

- Fear of failure or fear of the unknown and the possible consequences of a negative ROI. Fear that ROI will be a performance evaluation tool instead of a process improvement tool. Also stirs up the traditional fear of change. (Very much as seen with TQL in mid-90s),

- Successful ROI implementation requires much planning and a disciplined approach to keeping the process on track. Only a carefully planned implementation will be successful,

- False assumptions include: The impact of training cannot be accurately calculated; managers may not want to see the results of training and development expressed in monetary terms; the training processes a complex but necessary activity. Therefore, it should not be subjected to an accountability process (15).

As previously noted, all of the observations above are routinely encountered in either the application or implementation of programs or processes within the military. The fact that these observations are common to any organization should be a source of encouragement for the military. Even more importantly, because the accomplishment of often difficult and ill-defined missions is expected within the military, the officers and civilians within DoD have routinely overcome any of the challenges that are a result of the above.

Phillips' observations really address the culture of an organization, which is the bedrock of the military. Not all Services have stirring advertisements like those used by

the U.S. Marines that so clearly define the "warrior ethos" that is the basis for any of the Services, but each has a culture reliant on the warrior ethos. Further, as a result of the Goldwater-Nichols act as well as warfighting experience since the mid 1980s, the Services have a Joint culture, which compels each to man, train, and equip their forces in a manner that permits immediate interoperability (57). A review of many of the current books, which enjoy popularity in industry, such as Leadership Secrets of Attila the Hun, Patton on Leadership; Strategic Lessons for Corporate Warfare, and Wisdom of the Generals; How to Triumph in Business and in Life, derive their tenets of leadership, management, and business practices from military tacticians or practioners (58), (59), (60). It would seem logical then to use culture as a unifying force, agreeing that differences in Services or missions, in the case of the COCOMS, are minimal next to the cultural issues faced by other sectors.

The above raises the necessary question, how does culture apply to ROI? There are several key elements, recognized by industry and discussed by Schein, which have deep significance to the military. These very directly apply to the development of methodologies that address training in terms of costs, assessment, and ROI:

- Culture is the most stable part of any given organization. Culture change should, therefore, only be undertaken with full knowledge of how difficult, expensive, and time-consuming that will be,

- Culture assessment for its own sake down not work because culture is simply too complex,

- Culture forms around common experience. Any group or organization that has shared sufficient common experience will have evolved a culture in the sense of shared language, norms of behavior, common values, and assumptions about how things are and should be,

- It is the fundamental role of top leadership to manage cultural evolution. If the leaders do nothing, the culture will evolve by its own dynamics in response to environmental and internal forces,

- It is the specific role of trainers to educate the organization on the realities of culture and to show how culture relates to various change agendas that the organization may be struggling with (43).

The recognition of the role of culture in military organizations, especially in SFHQ organizations should permit the construction of measures for any area of concern to be nothing more than common sense. The need for preparedness to minimize casualties and to optimize the advantages of armed forces naturally drives any level of the military to assess performance, so assessment is part of the culture. The need to guarantee the taxpayers "value for their money" also compels staffs to develop cost and performance methods that justify the actions of each Service or COCOM. It is also a certainty that in the absence of direction, the need to be able to show readiness or efficiency is sufficient for even junior members to have some sort of method to display progress. Thus, each of the statements above, with the possible exception of the last one concerning trainers, is already inherent in the military, even in a Joint context. What is lacking is a uniform way to use the common culture of the Services to properly demonstrate ROI from the resources provided for training.

A recent study conducted by the Army Research Institute (ARI) investigated, "employing structured simulation-based exercises to explore and assess the methods, tools, and measures necessary to facilitate distributed wargaming. Key aspects of the environment were structured wargaming exercises for distributed staff and command groups, collaborative tools, and a set of performance measures to assess wargaming (14)." The environment studied had many similarities to the architecture, scenario basis, and objectives of the Navy's FST program. Further, the research addressed objectives, which are essential to the success of the FST program as an operational training capability, such as the need for a network supporting information displays, and communications systems representing the full spectrum of capabilities used in operational theaters. Two other specific objectives that are in concert with the requirements necessary to evaluate, assess, and certify a CJTF or FCC staff are:

- Design and develop structured exercises to support the conduct of planning and wargaming in a simultaneous, collective, multi-echelon and distributed manner. The distributed exercises require planning between higher and lower echelons (vertical integration) and across the same echelon level (horizontal integration). Tailor the design to an audience in which three to eight participants located in at least three separate or distributed locations interact directly and accomplish tasks collectively. Include collaborative, interdependent tasks (i.e., the task requirements for each participant will depend on the work of the other participants).

- Design and develop measures to assess the effectiveness of distributed wargaming. Performance assessment must address the outcomes of distributed wargaming to identify whether participants successfully identify and address problems or conflicts within a COA. Conduct a formative evaluation to gather participant feedback on the research environment, particularly the wargaming exercises and measures developed (14).

A challenge faced by the authors of the report was the fact that previous synthetic exercises had not utilized extensive horizontal or vertical elements. Though the horizontal and vertical elements were not necessarily viewed at the operational Level, the principle is valid and is a key consideration in assessing performance, cost, and ROI at the FCC and CJTF staff level. The observations of the participants in the investigation did not find horizontal engagement to be particularly difficult but found vertical engagement to be a challenge. At the operational level, horizontal engagement may indeed prove challenging since distributed simulations typically utilize on-site representatives with pre-scripted responses to portray the reactions of another FCC. In it's ultimate configuration, a distributed simulation would permit actual FCC staffs to interact with each other as well as with the CJTF staff. The observations of the participants in the ARI investigation are germane for both interactions at the operational level. The observations, which require incorporation into the current project, are:

- Participants in the vertical condition reported that rarely would subordinate commanders have the opportunity to participate in wargaming,

- Participants in the vertical condition stated commander-centered wargaming seemed more like mission rehearsal,

- Individual differences in wargaming experience undoubtedly affect performance. Future research might use background information, from demographic surveys, for example, to assign participants to duty positions and roles (14)

The summary of the ARI report noted that there was a need for more research and training permitting the Army to train in a commander-centered wargaming across echelons to meet its Future Force objectives distributed operations (14).

The synthetic training carried out by the Navy, Air Force, and under development by the Army permits all three Services to train with or at least interface with FCC staffs, thereby enhancing and improving the relevance of synthetic training. This capability cannot be truly viable until a method to assess the effectiveness of the staffs involved, the costs involved, and the ROI gained from the training can be reliably reported. Certainly, the Services derive no long-term individual benefit by supporting such training, even if it could be supported without jeopardizing individual, small unit, or deployable unit training objectives.

While the above are not unique, the questions about resource expenditure, the duplication of expense (which should be managed as a result of both the Quadrennial Defense Review (QDR) and the National Defense Panel (NDP), and the requirement to develop effective and timely training in a military context are obviously being debated in other sectors as well. Many decisions regarding future system acquisition and even exercise execution are being discussed in terms of all three of Kirkpatrick's points, even though they are not always viewed as the basis for training evaluation.

Though evaluating any type of training exercise would provide for the ability to analyze the points raised by Kirkpatrick, a focused look as to how a completely distributed synthetic training event could be the basis for continuing not only training, but certification and proficiency for either a CJTF or FCC staff would seem prudent. This would then lead to the ability to determine methods to improve future training events. If the opportunity to incorporate live elements in the training could also be evaluated in terms of performance enhancement or cost, then the ability to quantify ROI would also be

achieved. Due to the control of scenario, training audience, opposition forces, and time of play, a distributed training system would seem to offer the greatest opportunity for performance assessment, but also utility of the systems involved to achieve desired training go.

The need to assess the performance of the CJTF or FCC staff regardless of whether the exercise is live or synthetic could not be more compelling. Though the cost statistic stated by Phillips may be based upon industrial, commercial, or educational experience, it would be hard to believe that the value would be exceeded. In fact, owing to the fact that the military is familiar with evaluation, it would appear that this cost figure could be substantially reduced. Regardless, the need to couple the assessment of performance with cost is widely accepted as a pre-requisite for ROI in other sectors and should be addressed by the military.

Phillips provides a preview of a ROI model that lays out one possible methodology to calculate ROI (16). The most difficult step is the conversion of data into monetary benefits, isolating the effects of training and identifying intangible measures. One particularly interesting note in his discussion is what he terms "soft data," which is typically either subjectively or behaviorally based. Unfortunately, many key areas of performance of a CJTF or FCC staff which must be assessed and included in any attempt to quantify ROI fall into this category. The expansion of the metrics from sources such as JMETs, NMETs, or some other critical skill sets could greatly reduce the amount of soft data collected. Some examples of subjective data are those concerning appropriateness of reports, completeness of reports, quality of intelligence, and the reactions of subordinate commanders to information. These quantities cannot be measured by merely using a stopwatch, and even if time can be measured, that alone is often an insufficient measure of response or comprehension. As discussed in an earlier section of this proposal, it is very possible, and in fact desirable, to construct measures which depict some of the items above in a fashion that is quantitative, vice being merely "binary" in nature which provides little insight to the successes or failures observed in a staff' primary activities.

2.3.4 Contribution to the Body of Knowledge Regarding ROI

The entire focus of this project is to develop a methodology to determine Return-on-Investment (ROI) for training in the military context regardless of the scenario, live, virtual, or constructive. An obvious benefit of the research would be development of a method to directly compare the advantages of various training environments, specifically live vs. synthetic. Just the concept of what ROI means in this or any military context must be defined for the discussion to be relevant and useful. Certainly the pressure to limit budgets exists and maximizing productivity to increase profits is readily understood, but what is the product of a military organization? The ultimate product is the ability to conduct successful combat operations, which protect the interests, citizenry, and property of the United States. Unfortunately, the only true measure of this product is actual engagement (or perhaps sustained maintenance of a lack of combat in a nuclear arms context) in combat, with the metrics being the number of casualties and some form of victory (34). The next best measure is the readiness to carry out those types of combat operations, gauged by assessing realistic combat training. Thus ROI must center not only on the costs to produce successful combat training, but must necessarily account for methods to portray improvement in readiness, methods to view changes in tactics, techniques, or procedures (TTP), methods to assess the success of remediation or real time scenario improvements, and perhaps even the utility of new equipments. None of these are really traditional views of ROI as discussed in business applications.

The important questions requiring understanding and resolution are: What is the purpose of any model(s) developed? Is it for decision support? Could it assist in determining the best allocation for scarce resources? Might it provide gross insights into overarching problems in training or perhaps even actual operations? Should the model be a platform for strategic dialogue?

The answer to all of the above in a military training context is yes. In the context proposed, an effectively designed model utilizing not only accurate costing but also performance assessment should be able to provide insight and utility to all of these

questions, which are of critical concern to both the civilian and uniformed leadership of the U.S. Armed Forces as they are to any corporate executive board.

As to the issue of determining the best allocation of scarce resources, the underlying questions that must be answered are: what form of training is most suitable, does it provide the requisite training to provide operational readiness, and can the training proposed do so in the time allotted? The goal of this research is to provide insights into the three underlying questions by utilizing model(s) that accurately account for all costs and the performance standards achieved. The ability to view the training successfully achieved and the costs involved, or even training with notable but specific deficiencies could then be viewed in terms of what should be included in the training, what might be excluded, what can be deferred or replaced by another type of training, or modifications to the training process which will allow it to remain within budgetary constraints. Indeed, it should also be possible to forecast what the required costs for training should be and if the previously forecast funding is insufficient to permit the requisite operational readiness due to changes in the scope of required training; to justify the additional expense or show the areas where risks due to shortfalls in readiness could be anticipated.

From a strategic perspective, the use of both performance assessment and costs to craft a framework for ROI: permits long term planning, a more certain method of determining the probable outcome of a specific training regime or scenario and thus a way to forecast operational readiness and budget for it.

One of the most common methods for applying ROI is as a tool to predict performance as a result of an anticipated or projected investment in training, hardware (or perhaps software acquisition), or expansion of a business venture. Though there are some analogies in a military context, it is proposed that the ROI formulations developed in the course of this research could be applied in at least three instances in a military training/readiness context.

A priori, ROI should be a consideration when predicting the time to achieve a standard of performance if conditions from a previous training event of similar construct have been significantly altered. Another change of condition would be if JMETs were being used to assess the value of some experimental technology which should enhance training value, decrease the number of personnel required, or shorten the time required to achieve a standard. Finally, is a major recurring deficiency affecting either a mission critical JMET or some group of JMETs can be overcome by application of a new training regime, some different procedure, or perhaps a new technology, then the improved readiness (which can be measured) or even the addition of readiness will be apparent. If the changes discussed also permit either a reduction in the cost to perform the training or a decrease in the time required achieving operational readiness, again the ROI can be characterized.

If on the other hand, the new training regime, procedure, or technology does not seem to address either a significant JMET, or a sufficient grouping of JMETS for the cost predicted to implement use, then an objective decision to reject use is available. In the case where a new technology was utilized that was specifically introduced to improve performance, the proponent of the technology (who could be a support contractor, developmental organization, or even supporting training organization) would be supplied with the standards that were not achieved as well as a sense of the degree of the shortfall so that any additional costs for development could be viewed in terms of what had been expended to date and what further expenditures would be required to achieve an acceptable level of performance. A second instance where ROI can be realized is during the training event itself. If, because of the application of standards to measure capabilities, a unit has exceeded the required operational readiness goals, a discussion between the commander of that organization and the training organization can determine the value of continued training. This capability does exist in the present assessment systems, but since different operational level staffs, and more specifically, the boards, centers, cells, and bureaus within these staffs, learn at differing rates, the ability to characterize performance in a numerical method, especially in a four tiered schema, provides greater flexibility in training. Perhaps more advanced training which is purely to

hone already demonstrated skills, can be undertaken. Alternatively, perhaps the time prior to operational employment can be more effectively utilized at other tasks, so training can be terminated. In either case, savings have been achieved and these can be characterized in both time and monetary terms. (The savings associated with performing additional training of an advanced nature could be larger numbers of personnel available for specific tasks or the development of additional skills which can be applied in possible operational scenarios relieving the need to augment the deployed staff, or by having procedures pre-written which will save time and permit commencing operations earlier than is normally expected. Again, not traditional ROI, but certainly examples, which are readily appreciated in the often-dynamic situations encountered in a forward theater.)

Finally, after the fact, could ROI be applied in a meaningful way using the models and methodologies developed in the course of this research? The unique nature of the types of training required and the need to improve it nearly continuously permit one more view of ROI. If a staff has not achieved the required operational readiness, the deficiencies can be easily traced to JMETS. Remediation for deficiencies can be specifically applied to the JMETS necessary and in a synthetic environment at a very readily determined cost. Thus, not only can successful remediation be demonstrated, the time required and the precise cost to accomplish the remediation is also available. The ROI principles that can be applied in this instance are the savings realized in comparison to similar live training, which may not be achievable or practical, the ability to portray the cost of remediation in terms of time and cost, as well as the ability to portray the remediation as a percentage of the original training event. The ability to demonstrate operational readiness in absolute (or nearly absolute terms) also assures the commander of the unit of that unit's readiness and those who are employing it of its ability to perform on arrival. Though somewhat less tangible than the previous examples, the latter statements are none-the-less very real concerns that can become operationally limiting.

Another example of after-the-fact ROI would be the savings in time, reductions in personnel, or even reduced costs for travel which can be achieved through the alteration of a deficient scenario, especially in a synthetic training event. A change in a scenario

and thus the sequence of events to be observed permit accomplishment of training in a more efficient manner or permit the elimination of ambiguity which led to t errors made by the training audience which may have unnecessarily extended the time required to accomplish operational readiness. A slightly different view of this same issue might be to consider changes to the scenario or even the performance standards themselves as requested by the training audience once that group has arrived in the operational theater. They can state exactly what changes are required due to circumstances that have changed since the training was administered or that might not have been clearly understood at the time the training took place. In any event, the required changes can be directly related to JMETs, and those specific criteria also reflected in the costs (certainly in a synthetic environment) required implementing changes. As a result, the time for the next training event will be more effectively utilized, and either a shortened training period or the additional costs (in terms of augmentation, additional assets, or delays in arrival) will be avoided. These can then be characterized as ROI as a percentage of the costs of the original training.

The unique goal of this research effort is to develop feasible methodologies that can be applied to existing or formulated processes used by the Services, COCOMS, or major headquarters to portray costs for training regardless of the environment utilized. The examples discussed in the preceding paragraphs and the reliance on both performance achievement and costs are an initial step in a consistent approach for portraying ROI.

Admittedly much of what is proposed in terms of ROI formulation can be considered an application of existing capabilities as contrasted with a new theoretical approach. The unique feature of this research is the use of a single basis, the UJTL, for both performance and costing. This approach has not been attempted to portray the results of performance in relation to resources expended and then to calculate ROI. As was mentioned in Section 2.1.4, the combination of hard and soft measures is also uncommon, and using this approach at the Operational Level of War (OLW) can create potential insights into readiness and resource allocation that are not currently available.

The synthesis of the assessment and formal costing methods as well as the fact that this research can be carried out regardless of the type of training or operational environment and regardless of the scale of the scenario utilized that is innovative.

The preceding discussion proposes the possibility that ROI does have several alternative formulations. It also hypothesizes that these are quantitative in nature, and in the final calculation, ROI will still be merely some ratio of cost, capability, time, or procedural improvement. The underlying pillars of performance assessment and costing are the most obvious combination of terms utilizing the common operational template comprised of JMETs that must be related to provide the basis for ROI.

In any other formulation, ROI can still be viewed as related to achieving a pre-determined or previously achieved standard of performance, but that performance could just as easily be based in numbers of personnel required to achieve the performance. A change in the numbers of personnel to achieve an acceptable level of performance is also a form of ROI since a reduction would indicate an improving capability, and thus a positive ROI. Using an established standard of performance, in the military case, an operationally ready state, permits formulating ROI in terms of changing procedures which could reduce the time required to achieve readiness, changing equipment suites, which could reduce not only the total amount of equipment required to achieve readiness, but perhaps also reduce personnel and time. The ability to view the success or failure of a change of condition and the subsequent performance thus establishes a more objective means of evaluating proposed personnel, equipment, or procedural changes, and the advantage or disadvantage of those proposals. Since the costs of these changes may also be available, there is still, if desired, an ability to demonstrate a monetary linkage to ROI. The ability to apply these various formulations of ROI at several junctures in the training process can provide insight, decision-making capability, and the ability to target the allocation of resources, and ultimately guide the strategic operational readiness goals of the U.S. Armed Forces.

3 RESEARCH METHODOLOGY

This chapter provides a description of the research completed in developing a repeatable and reliable methodology to determine Return-on-Investment (ROI) for an Operational level staff regardless of training environment. The initial discussion will center on defining the research problem to be addressed concerning the current capabilities to quantify ROI in a military context at the Operational Level. The necessity of considering both common method of performance assessment and costing models will be elaborated on. The subsequent section will describe the methodology and technical approach utilized in constructing a more useful and more practical method of quantifying both performance assessment and costing based upon the Universal Joint Task Listing (UJTL) (1). The specific steps required to complete structuring this model will be described herein. A presentation of the experimental processes that will validate the methods proposed and the analyses required in determining success for this research will follow the discussion of the methodology development. Finally, potential applications and the limits of this research effort will be offered for consideration.

3.1 Research problem

There is no current reliable or repeatable method to portray ROI for training in any large-scale environment whether Live, Virtual, or Constructive (LVC). This research focuses on developing and demonstrating the feasibility for new formulations of ROI based upon performance assessment and costing. The ability to view ROI as it applies to specific training scenarios, objectives, or operational readiness will lead to improvements in U.S. military operational readiness, training effectiveness, and potentially lead to more uniform approaches to the acquisition of training technologies and infrastructure.

3.1.1 Problem Definition

The U.S. military, and more specifically those commands such as Joint Forces Command (JFCOM) have struggled with the need to achieve the very high state of operational readiness required in the current high-tempo operational environment, primarily utilizing the exploding capabilities provided by synthetic training, while demonstrating a tangible ROI in relation to the very large investments that have been

made in the Joint National Training Capability (JNTC). Any enterprise, regardless of size, concerns itself with balance that must be struck to ensure maximum productivity and the investments (which detract from profits) required to assure this productivity. The area of training is difficult to characterize in an industrial environment even though there are some, fledgling (in some instances) methods of determining the Return-on-Investment (ROI) for a specific investment in personnel, training, or production. This is not as readily apparent in a service related industry, and is extremely difficult, at least with conventional formulations in a military organization. More specifically, the ability to quantitatively demonstrate the effectiveness of training received in a modeling and simulation environment is not readily available (61). Yet, it is entirely reasonable to ask, what is being gained from the very large investments that have been made in the Joint National Training Capability (JNTC) as well as by the individual Service's and Combatant Commanders (COCOMs) in simulation. As early as 1995, DoD recognized the power and promise of simulation and stated that,

> " M&S will enhance readiness by allowing UCCs[14] and Services to train forces, develop doctrine and tactics, assess performance of units, support planning, execution, and analysis of operations and exercises, evaluate operational plans, conduct "what if" analyses on those plans, rehearse missions . . (62)"

Unfortunately, the assertion by Ogus is still true; the goal of this proposal is to devise a method to reliably demonstrate the return on investment gained from modeling and simulation.

3.2 **Research Approach**

3.2.1 Overview

As described in the preceding sections, the U.S. military requires a method to reliably demonstrate ROI for the substantial investments that have been made in simulation systems and in the JNTC. Though there are several possible levels where

---

[14] UCC – Unified Combatant Commands

solving this issue apply, the ability to relate the tactical level to the strategic level utilizing the operational requirements and capabilities inherent in either a CJTF or FCC staff provides the greatest common gain for the COCOMs, Services, potential Operational level staffs, and the trainers. The operational level is the first one where Service background either begins to lose importance, or is nearly irrelevant. As such, having a feasible methodology to portray readiness, costs associated with attaining readiness, and finally the ROI associated with both of these pillars appears to be a logical and achievable goal.

The methodology developed in the subsequent sections of this dissertation delineates the specific processes, software tools, mathematical basis, and finally some formulations for ROI that can characterize the benefits of current expenditures in terms of readiness and training.

As described in previous sections of this dissertation, the ability to devise various formulations of ROI to arrive at one that is most applicable for a specific situation depends heavily upon performance assessment and costing methods. Therefore, a feasible methodology for performance assessment that can be readily used by either the senior leadership of an operational level staff or an outside observer team must be constructed. This methodology must apply to a military context and more specifically to the correct level of operations. The Operational Level (OP) JMETL and the subordinate tasks and measures form the basis for constructing the methodology. The approach developed in this dissertation follows the guidelines already set forth in the Joint Training Manual for the Armed Forces of the United States (8) in that standards for measures will be quantified, but the method used to combine measures to provide a "warfighter's view" of readiness, will be an extension of the method proposed by Startin, described in the background (23). This methodological approach is specifically designed to measure tasks within the appropriate portion of the staff, for the correct mission, and then to permit the evaluation of effectiveness of the required task on staff performance in support of the commander. Although the implementing instructions governing training and the UJTL have been existence for more than five years, the movement from qualitative measures or

strictly binary measures to evaluate performance has lagged. This dissertation is a first step towards a implementing a more objective framework that can be utilized by any major staff with measures that are appropriate across the Range of Military Operations (ROMO), in either a CJTF or FCC role. A second focus of this development will be a formal method for cost accounting that can directly relate to the JMETs used in training. A uniform method of accounting for costs which permits direct correlation to training measures, staff processes, or JMETs, will also provide a vehicle for major staffs to predict costs for a specific mission or capability, something that is currently attempted, but based almost exclusively on either past events or an informal estimation of what a specific mission might require contrasted with available capabilities.

Some key issues that will have a bearing on the methodology development and validation in subsequent sections of this dissertation are:

- The development, testing, and initial implementation of the methodologies and related tools will be done in an unclassified format. A great deal of data, current processes, and the primary documentation is unclassified and though there is information residing in the classified realm, the efforts here will permit greater access, collaboration, and usage if they repose in an unclassified format,

- The vast majority of the methodologies and tools herein are already in existence, such as the EXCEL application. The relational database tool RATIONAL was evaluated for applicability. The impetus for these choices was to utilize applications which are either well understood or already funded but not fully developed such as RATIONAL, but would seem to provide a very useful platform for this research (63)

- The role of operational level staffs is such that they may not derive substantial, if indeed, any real benefit, especially in a distributed synthetic environment for training, from the use of live participants at the tactical level. Either semi-automated or computer-generated forces appear to provide a more robust response and more flexible environment in which to assess the complex problems these staffs must overcome, especially in constructing a suitable and realistic opposition force,

- Regardless of the specific armed service from which the Commander and the majority of any operational level staff are comprised, the nature of tasks which are required for the staff in a CJTF context tend to be driven by the theater requirements. Service experience for the Commander and the staff add perspective, but the challenges that must be overcome are common to all CJTF staffs, and thus using the UJTL as the basis for performance assessment and costing should be applicable to any operational level staff.

Validation of the methodologies proposed would be conducted in both a simulation environment and in environments where either a CJTF or FCC staff is undergoing training. Both these venues will provide opportunity for statistical verification of the methods chosen as well as the utilizing the knowledge of Subject Matter Experts (SMEs) to verify the applicability of measures, calculations, and results.

3.2.2 Methodology and Technical Approach

The development of a framework that will permit various formulations of ROI applicable to training improvement, scenario development, cost reductions, and technology evaluation, requires the development of both a quantitative methodology to assess staff performance and a costing methodology that can be linked to specific measures, processes, pillars, functions, JMETs, or if absolutely necessary events.

3.2.2.1 Performance Assessment Methodology

The overall goal of the Operational Level staff is to develop plans that fill the strategic needs of the COCOM, integrate the needs of the FCC staffs, and direct the actions of the subordinate tactical commanders. These goals are portrayed in the figure below and are often described as the Assess, Plan, and Execute (APE) model. Included in the Execute phase is the issue or execution of the orders that implement the plan as well as monitoring the responses and actions carried out due to transmitting the order. Assessment refers to assessing the effects of the plan as executed and adjusting the plan to obtain the desired tactical, operational, and strategic effects.

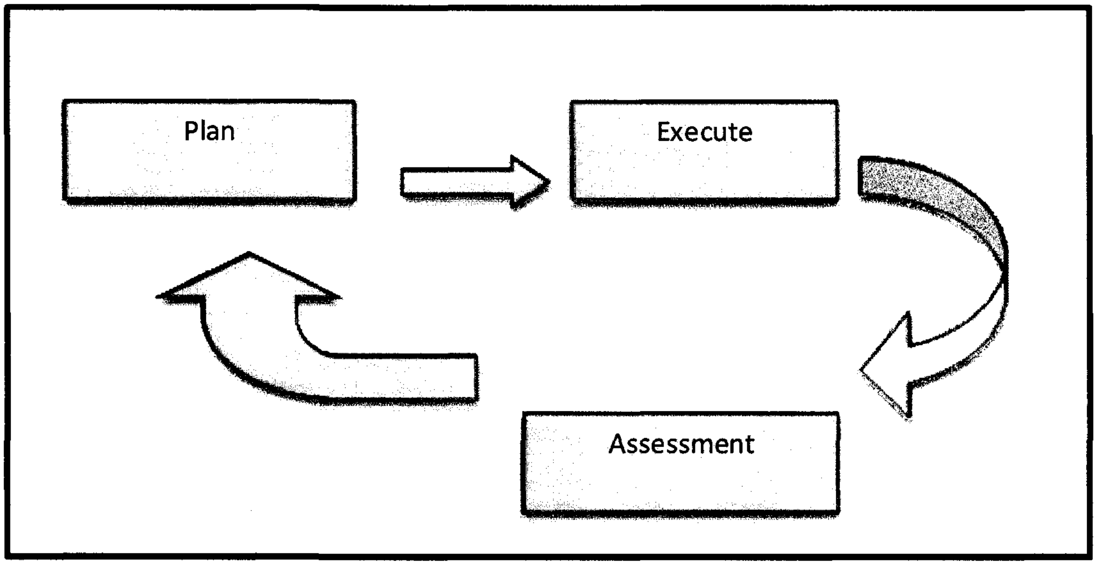

Figure 9: Staff Actions Functional Construct

The best source for quantitative measures exists within the UJTL applications already being used by the COCOMs, existing JTFs, and Service Functional Component Commanders. While some of these are held in the classified realm, there are ample sources available for this work.

As noted in the overview, Service culture tends to bias the measures that might be most useful for a JTF staff, but it provides an excellent basis for those required for a FCC staff. This dissertation utilizes as its primary sources the measures developed and either in current use or those being used for training and certification events by the Standing Joint Force Headquarters (SJFHQ), Commander Second Fleet (C2F), or in the Blue Flag (BF) exercises for the U.S. Air Force.

To provide the most robust coverage of measures, which also provides the greatest possible application to the probable roles or missions that could be required of any CJTF or FCC staff, all the currently existing measures will be used for performance assessment. If additional measures become available (from a source such as Pacific Command (PACOM) or another source) these additional measures will be incorporated as described herein.

Not surprisingly, there is a degree of overlap already existing in the JMETs and measures, as well as in a substantial number of cases, specific measures. This is attributed to the fact that all the organizations named are functioning within the hypothesis regarding the operations requirements at the Operational Level.

There are 2530 measures associated with the seven Operational level areas; the three organizations mentioned have added 212 experimental measures to more accurately portray the measures required to assess the performance of some of the JMETs. Of this enhanced set, 972 are currently either actively used or being considered for use by the three organizations cited. Slightly less than 10%, 105 have been considered by more than one organization, and of these 30 have significant differences[15]. These differences stem from various causes, some are merely the choice of differing units from the suggested standard, some appear to be attributable to the differences in culture, and some seem to be due to possibly a difference in the perspectives of the organizations that originated the standards.

The first issue requiring resolution is not only establishing a single value for the standards in question but also having the originators consider the values suggested by at least one other organization. There was some consideration early in the research effort to proposing at least two courses of action to resolve the differences in standards. One was a mathematical resolution; the other was simply to delete the values from consideration if some agreement on the value could not be reached. Neither of these courses of actions was required since the three organizations involved agreed to use values for the standards in question proposed by the researcher. An example of a set of values which were accepted was those for logistical aircraft normally controlled by the U.S. Air Force. The U.S. Navy agreed that the standards proposed by the U.S. Air Force were the most appropriate where logistical aircraft measures were concerned.

---

[15] A significant difference is considered to be more than a 25% difference in the value of the standard considered to be sufficient to attain certification.

The second step leading to a meaningful method of quantifying the staff performance is to decompose measures from specific JMETs by either JMET area or staff process, which will relate to the organizational elements contributing to the major functions. This concept was introduced by Startin when investigating Warfare Commander performance at the tactical level. Startin discussed the fact that assessment using the Navy Mission Essential Tasks (NMETs) required modification in order to be capable of producing accurate assessments by automated means. To do so required measures to be categorized according to skill type and these essential skill categories incorporated into the METL paradigm (23). Applying the logic Startin utilized to organize measures into appropriate JMET groupings, so that the measures more accurately described by sets of skills requiring proficiency at the operational level is one way of ensuring that the measures are correctly grouped for use in later steps. Another check of measures is to relate them by using a development conceived of by C2F. The process model devised by C2F is similar to the method developed by the U.S. Air Force in developing the model for an Air Operations Center (AOC) that detailed 51 key processes. The C2F model details 59 core processes that lead to the organizational elements (or cells), which ultimately supply the APE model. EXCEL™ had the functionality required to organize the entire JMETL into groupings which represented the C2F core processes. When grouping task was completed, measures were identified which were necessary to complete any of the specified processes.

Although organizing the JMETL into processes was useful in terms of verifying the measures necessary to accomplish the processes, the ability to assess processes still depended upon specific boards, centers, cells, or bureaus which were part of the Organizational Elements. The processes are not always carried out in a sequential fashion, and those performing the tasks within the processes are also not limited to completing tasks in a sequential fashion. These factors caused the researcher to organize the JMETL into either Organizational Elements (OE) or APE groupings to build the models used to assess performance in trial runs and for actual observations.

Validation of the choices of JMETs, measures, and the appropriate grouping was carried out by consulting several Subject Matter Experts (SMEs) who either routinely execute large scale exercises or who were involved in the actual assessment of an operational level staff during a major exercise.

Two methods of validation of the process to select measures that are appropriately grouped into the organizational elements, either by way of core processes or JMET grouping[16] were available. The first method to validate the measures was afforded by the availability of two different teams composed of SMEs in different locations who were working on a project that also required a review of JMETs necessary for an Operational Level staff to perform in various roles over the range of military operations.

A comparison of the agreement between these two groups regarding the assignment of JMETs to either Organizational Elements or to APE groups was completed. prior to visiting all the major fleet headquarters staffs. A second review was conducted by each Numbered Fleet staff when assessing the applicability of the Organizational Elements, APE functional areas, and JMETs to each of their required operational capabilities. These reviews were incorporated in the statistical analysis of the JMETs and Organizational Element groupings.[17] As an aside, during all the aforementioned reviews, the accuracy or at least applicability or validity (e.g. Does this measure make sense and is it a standard that can even be attained?) of many of the measures was expected. Any data that could be captured which was used for assessing actual events was included in model formulation and in the analysis of the overall results

---

[16] In many cases the JMET grouping and core processes bear the same name and description, so only those processes which are unique or group JMETs uniquely need be considered in the mapping from measures to organizational elements.

[17] Though the Numbered Fleets reviewed the Organizational Elements, only C2F had the time and certification event requirements which supported quantitativel comments on the accuracy of the JMETS, Organizational Element, and APE assignments made during the course of this research

	Military Engagement, Security Operations and Deterrence	Crisis Response and Limited Contingency Operations	Major Combat Operations
FCC			
JTF			

Table 2: Roles and Range of Military Operations (ROMO) (2)

The next step is to weight each of the processes according to the role and military operation being undertaken. Table 2 shows the possible combinations to be considered in this research.

The role and, perhaps more importantly, the military operation in question determines the level of importance that any individual process will assume. For example, the process "command forces" applies in any of the above, whereas the process to "develop targets" has diminished importance in Peace Operations, and perhaps has no place in Foreign Humanitarian Assistance operations. Further, there may be a very minimal role for a FCC to "command forces" in a Peace Operations scenario, while the majority of the actions required within this process would be carried out by the CJTF.

It is also possible that even within a specific process, the measures that apply from any JMET may be less important dependent upon the role and mission than in another instance. The mapping of measures to processes, and the weighting of processes for a given role and mission will require review by the organizations cited, as well as numerical testing to determine the impacts of the assigned weights on overall performance.

With the measures mapped into appropriate processes, the next step is to identify which measures apply to the organizational elements that contribute to daily operations. These organizational elements are portrayed in Figure 10

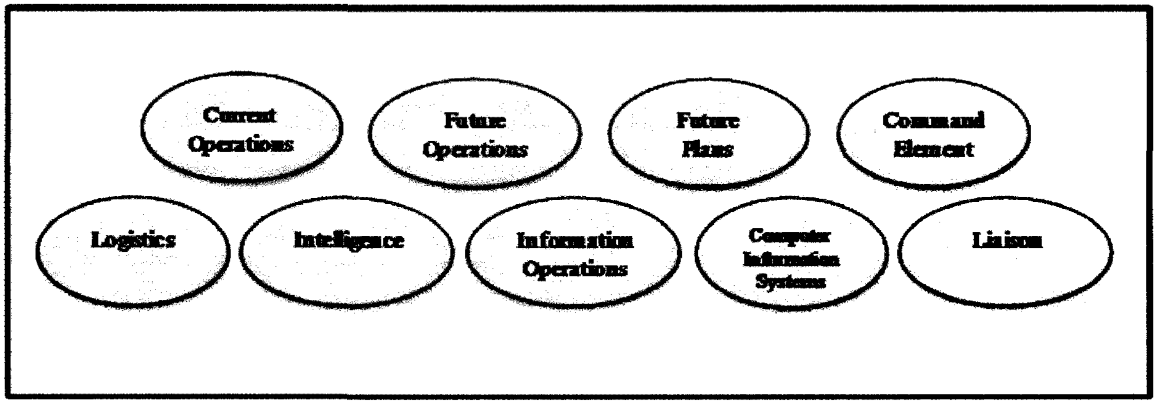

Figure 10: Key Organizational Elements

The method of applying the measures that are most appropriate is directly taken from the process selection in the previous step. Utilizing the insights gained from mapping measures to processes, it is possible to create a logical mapping of measures into the elements shown above. Another delineation that must be made is the method of distinguishing between Current Operations (COPS), Future Operations (FOPS), Future Plans, Command Element, Logistics, Intelligence, Information Operations, Computer Information Systems, and Liaison.

The primary determination for COPS, FOPS, and Future Plans is time frame. Anything occurring within a 24 hour period will be considered COPS, 24 – 72 hours FOPS, and beyond that Future Plans. Decisions or actions required by the Commander or actions regarding connectivity or the monitoring of communications will also be grouped within the Command Element pillar. Those measures that do not directly affect COPS, FOPS, or Future Plans, but relate to Logistics, Intelligence, or Information Operations will be collected in the appropriate OE.

Now that the measures are "binned" in the appropriate element, they will be averaged to arrive at an overall measure for each pillar. The averaging method proposed by Startin was introduced in Section 2.1; a specific example of how it will be applied during the research for this dissertation follows:

Since the ultimate goal of the process is to relate measures to either OE or the APE MFAs, and these will consist of combinations of LDFs, the skill values that make up the factors for the LDFs are the term being generated in equation 3.1. There will be another weighting factor required later in the assessment calculation, but the weighting factor above deals with specific measures in relation to the "skill" bins necessary to form the LDFs (23).

$$S_i(T_j) = \frac{\sum w_j T_j}{\sum w_k} \tag{3.1}$$

It is also important to note that for either the OE or APE MFAs, not all "skill" bins will apply; only those that apply to a specific OE or MFA are used in the calculation. It is also possible that within skill bins, specific measures may change weight, as might be true for the OE depending upon the role and operational context within the ROMO that is being undertaken by a particular Operational level staff. A good example are the measures which relate to Foreign Humanitarian Assistance which would have little to no importance during Major Combat Operations, but would be the focus of effort during Foreign Humanitarian Assistance operations, whereas some of the measures relating to time sensitive targeting or battle damage assessment are irrelevant during Foreign Humanitarian Assistance operations. Instead of changing the values for the standards, a more realistic method to view necessary measures is to change the weighting to reflect the operational imperatives at hand.

To illustrate the result of equation for a specific skill bin, a set of measures, which will eventually relate to the Command Element OE, a specific skill, and the relative importance of each skill (weight) will be combined as set forth in equation 3.1.

Leadership	Trust	Confidence	Balance
Competency	Authority	Balance	Decisiveness
Initiative	Responsibility	Accountability	Courage
Situational Awareness	Adaptability	Projection	Objectivity
Negotiation	Accuracy	Clear Communication	Preparedness

Table 3: Potential Skills and Competencies

Table 3 shows some of the skills that are relevant to either the OE or APE MFA constructs. There may well be additional skills or competencies required to fully define the capabilities required to characterize all OEs or APE MFAs accurately, but this listing provides and excellent basis. It is noteworthy that Table 3 is a combination of three different views of competencies, one from the operational experience of the researcher himself, as well as from both Dr. Winston Bennett and Dr. Will Startin (38), (23), and that the lists of all three were very similar.

The skill area relating to the Command Element for this illustrative example is accuracy. The relevant measures from the UJTL are shown in Table 4 below with their certification standard, a hypothetical observed value, and a hypothetical weighting. An example of the weighted average score for a given skill assignment, such as those measures relating to accuracy in Table 4 and Equation 3.1 results in an overall skill evaluation for the accuracy skill of 0.91. The hypothetical observed values would support this, and it would not be an unexpected result for a staff at the beginning of a

certification exercise or after a period where the staff had not operated in the Operational Staff mode.

Measure	*Standard*	*Observed Value*	*Weight*
Accuracy of information in plans and orders issued and disseminated to subordinate units.	95 – 99%	90%	.9
Accuracy of mission-essential information maintained on situation displays	90%	97%	.8
Accuracy of intelligence situation displays maintained with mission-essential information.	85%	90%	.8
Accuracy of data used by operations staff.	85%	85%	.8
Accuracy of deployment orders and notification requirements transmitted/disseminated within reporting criteria.	95 – 99%	97%	.6
Accuracy of data transmitted/disseminated.	95 – 99%	95%	.7
Accuracy of information in ROE changes.	95 - 99%	90%	.8
Accuracy of information in essential logistics, maintenance, aircraft, missile and personnel reports.	95 – 99%	95%	.5

Table 4: Measures, standards, observed values, and weights for the Accuracy Skill

The advantage of averaging the measures in this manner is that the measures are being averaged in relation to others that directly relate to the capability for each of the organizational pillars to impact the performance of the staff vice in the JMET groupings, and thus are measures of performance and not a rating of the JMET tasks.

The final step in arriving at some estimation of the capability of the staff's performance assessment is to aggregate the organizational pillars in a fashion that reflects human learning. Each of the major functional areas (MFAs), Plan, Execute, and Assess will be calculated as a product of the organizational pillars that are relevant to that functional area. The problem to be resolved here is how to correctly aggregate the pillars to characterize the functions correctly?

The overall goal is to start with a set of processes defined by known measures, and ultimately produce plans that can be issued as orders for action, collaboration, or concurrence both vertically and horizontally. The key inputs to each of the modules in the A, P, E construct are the organizational pillars which directly impact, to varying degrees, each of the major functions.

The proposed research will utilize two possible approaches to arriving at performance assessment using established measures, with the understanding that it would appear that the logarithmic approach originally proposed by Startin has the greatest potential for success.

The first approach is an extension of a method initially developed by McGinnis and Stone for unit training. This method is the most basic in that the mathematics involved is very simple, and that may not permit some of the complexities of operational level staffs to be accurately represented. The method may provide some high level information and possibly insights into the importance of some measures, Organizational Elements (OE), and perhaps even into the MFAs of the staffs, Assessment, Planning and Execution (APE) It should be noted that Execution, at the operational level involves issuing orders and monitoring the communications, compliance, vertical and horizontal information flow, and reactions to plans and orders. There is no expected delivery of weapons or direct tactical employment.

As already mentioned, the two "target" sets of data that need to be related to the measures are the Organizational Elements (OE), consisting of:

- Current Operations (COPS)
- Future Operations (FOPS)
- Future Plans (FP)
- Command Element (CMDE)
- Logistics (Log)
- Intelligence (Intel)
- Information Operations (IO)
- Computer Information Systems (CIS)
- Liaison (LIA)

These organizational elements will be used in all the approaches since they are the common framework that develops plans, products, or other information for one of the major functional areas.

Some issues universal to all three approaches are:

- What skills, attributes, or other defining characteristics can be used to logically "bin" the measures into the eight OE listed above?

- Though each OE has measures that apply, some of the measures may apply to more than one OE, how should these be handled?

- Though this initial treatment specifically delineates three of the elements, IO, CIS, and LIA these may not be significant enough on their own to influence or impact any of the approaches, and may need to be incorporated into the CMDE or COPS OE,

- Regardless of how many OE are ultimately used, there will need to be at least one, if not two applications of weighting measures. The

first would be to the chosen attributes so that these are properly represented when computing a value for each OE, and in at least two of the approaches, a weighting for the importance of the OE to the values for the APE functions,

- It is desirable to have all three approaches result in values for each of the APE blocks so that the validity of a specific approach or comparisons of approaches are logical.

Once the attributes are selected, the three approaches will proceed as described in the following paragraphs.

McGinnis and Stone originally proposed that aggregation of tasks in a team setting, to produce a performance evaluation in a synthetic environment could be accomplished by averaging and weighting at successive levels (8).

The algebraic approach will combine attributes that are appropriately "binned" to each of the OE. The bins will be averaged to provide a single value for the bin, and then weighting factors will be applied to the bins. The weighted values will be summed and then averaged to permit arrival at a single value for each OE.

The OE combination is similar in that each OE will be weighted as each applies to the APE construct. The applicable OE (not all apply to each, since measures for Logistics will only reflect those measures which are specifically required for planning and not execution. The same reasoning will apply to IO, Intel, CIS, and LIA, though these may apply to other areas, typically either Assess or Execute.) The weighted OE scores will then be averaged to arrive at a single value for APE.

The second approach consists of two alternative branches, each of which provides a quantitative measure for staff performance. The first alternative branch will be considered a two-level aggregation based upon the earlier work of Startin (23) that was discussed in detail in Section 2.1.1. As noted, this approach will require that the skills discussed in the averaging example be used in building the LDFs. In this construct, the attributes are again binned, they are also averaged as before, and the bins also require the

assignment of weights in relation to the importance each bin has relative to the OE "target" for which the calculation is being performed. As also described by in Section 2.1.1, normalizing factors will need to be determined so that values greater than 1 do not cause artificial or nonsensical scores after combination. Once these have been determined, the LDFs will be combined in the product relations described in Section 2.1.1 to arrive at values for each OE.

The final approach is a single-level aggregation. Once again, the attributes already selected for the previous methods will be used. In this case, there will be far more measures "binned" under the attributes since the "targets" for the bins are the APE functional values. As in the two-level aggregation, the bins will require the assignment of weighting values, and the determination of normalizing values. The mathematics for this approach are completed in a single step since the LDFs and normalizing factors lead directly to the APE values.

Since the last two approaches are logarithmic in nature, they should reflect the learning capabilities or decrements of the staff. Which approach will more appropriately portray those trends is not clear. All three approaches should provide insights into the attributes that are most important for staff operations, and, in two of the methods, the relative importance or impacts of the OE to the overall performance of the staff.

The three factors that will require research and refinement for all three proposed formulations are the skills, weightings, and normalizing factors. Skills will have attributes such as timeliness; accuracy, situational awareness, coordination, data fusion, discipline, initiative, and communications proficiency to name a few that will be required. Not all skills will relate to all the OE categories, but enough must be devised to accurately portray the knowledge, skills, and abilities (KSA) necessary to perform effectively. The U.S. Air Force has had quite an extensive body of work performed in this area, lead by Dr. Winston Bennett at Nellis Air Force Base. The paper he co-authored on Integrated Performance Measurement and Assessment in Distributed Mission Operations Environments: Relating Measures to Competencies (38)as well as a very thorough paper regarding Joint Force Air Component Commander (JFACC) Mission

Essential Competencies (37) will be used to help define the skill areas that will be used to "bin" measures for the OE categories.

The two remaining issues from above are inter-related, that is the weights and the normalizing factors. The relationships between these was discussed extensively in Section 2.1.1, and equations 2.4 clearly shows that weighting factors are the only variable used in determining the normalization factors. Therefore, not only to accurately portray the importance of various skill sets have to either organizational element performance or major functional area performance, but to ensure that the performance assessment values correctly portray staff capabilities, the determination of weighting values is the key area for development. There has been no previous research in this area, and one of the extensions of the Startin methodology is a mathematically based approach to determining weighting factors. The determination method proposed for this research is to conduct a regression analysis of weighting factors until little change is noted in either of the logarithmic methods discussed. The averaging method will be used to come up with an initial estimate of weighting factors. Once these are determined, then values for all measures will be randomly generated, and then averaged in the skill bins. The next step will be to predict the overall OE or major functional area grades based upon these skill scores and determine the weights required. Using an iterative approach, weights will continue to be estimated until the weights, which most appropriately represent the skill bins, are determined. These weights will then be used in conjunction with actual observed data to validate performance.

3.3 **Research Plan**

3.3.1 Research Steps

The following sections outline the sequence of steps completed in support of the research to arrive at formulating ROI for use by senior decision makers. The steps included the research and data gathering to determine the currently used or readily available performance measures and cost data, the selection of the most appropriate software applications for this task, the development of the algorithms necessary to represent costs, performance and ROI, the construction of the appropriate modules

necessary for computation, the simulation testing to validate the outcomes, the operational staff and observer evaluation of the methodology, and use of the methodology during operational level staff certification.

i. Determination of currently used or available cost and performance measures

a. Completed an in depth survey of metrics used by all the Service components and by JFCOM and other four star staffs to evaluate the performance of either FCC staffs or CJTF staffs. Any metrics that are used to certify or pre-test such staffs have been included to develop the most complete set of metrics possible;

b. Completed an in depth survey of the cost modeling tools used by all the Service components, by JFCOM and other four star staffs to estimate the costs involved with synthetic training at the Operational/Strategic level;

c. Completed an in depth survey of the cost modeling tools used by all the Service components, by JFCOM and other four star staffs to estimate the costs involved with live training at the Operational/Strategic level.

ii. Software Selection

a. Surveyed appropriate applications, database applications such as EXCEL™ as well as relational database applications are necessary to permit relating costs, measures, JMETs, and perhaps capabilities for the final ROI methodology. The RATIONAL REQUISITEPRO™ database was offered for use since it was funded at JWFC and had personnel who were available to support the research.

iii. Performance assessment algorithm development

a. Three approaches required three different mathematical formulations.

b. Skills which most appropriately represent the key warfighting Knowledge, Skills, and Abilities within either the Organizational Elements or Assess, Plan, and Execute functions were assigned to all measures for each JMET which applied to each model.

c. Determination of the appropriate weighting for the "bins" associated with the OE categories, and APE major functional areas was performed.

d. Determination of normalization factors to be used in the logarithmic driving functions for the two logarithmic variations was performed.

e. Selection of measures from JMETs that apply to either the OE categories or the APE major functional areas was performed in the instances where no measures were available to assess a necessary Mission Essential Task.

f. Spreadsheets were developed to calculate the values for performance which were identified in steps a through e.

iv. Costing assessment algorithm development

a. Cost modules were constructed for fixed/direct costs, personnel costs, indirect costs, and one time/specific event costs;

b. Costs in each category from paragraph 4.a. were related to the most descriptive and specific criteria possible starting with individual measures and proceeding to specific event if necessary;

c. Costs were compared on the basis of JMETs, providing potential insight into the utility of resource expenditure for individual JMETs as well as potentially allowing for the development of a scheme to identify costs that can be reduced by the infusion of technology or

process so that personnel, fixed/direct, or perhaps even indirect costs could be reduced in future events.

v. ROI algorithm development

a. ROI algorithms that portrayed either the ROI realized by conducting similarly based JMET events in a LVC or purely synthetic environment were developed. The ROI that could be realized in successive synthetic events was also calculated. The detail provided by both the J-9 staff and the Operational Level staff, permit using a similar approach for determining ROI in terms of time and cost predictions to achieve desired level of training, utility of additional training in either distributed mode or some alternative to complete certification or remediation of shortfalls, contrast of live vs. synthetic training, scenario selection, and technology improvement predictions.

vi. Test Methodology

a. Once devised, regardless of the model (Major Combat Operations, Noble Resolve, FA PANAMAX, or TOPOFF 04) being investigated, the spreadsheet mathematical functions had to be tested to ensure accuracy.

b. Various combinations of scoring distributions (in a range which had the specific values for T1 (95%), T2 (85%), T3 (75%), and T4(0%)) representing the researcher's knowledge of operational performance were applied to initial models to permit a baseline of possible performance to use as a comparison to actual performance.

c. Using appropriate measures, run a series of linked tests to validate improving and declining performance and the impacts of parameter variations (a "bad day" and the impacts on overall performance in either an entire OE category, or in a selected set of skills)

vii. The OE categories, measures, skills, weights, and LDF factors were reviewed for applicability by several subject matter experts from several organizations and operational level staffs.

a. The primary staffs conducting the review were Second Fleet, personnel from the J-9 Directorate of Joint Forces Command (JFCOM), and personnel from the synthetic training division of Tactical Training Group Atlantic.

b. Personnel from the Joint Training division of U.S. Fleet Forces Command also reviewed the measures, skills, weights, and LDF factors.

c. Vice Admiral Chanik was consulted as the Operational Level Commander regarding the applicability of the approach and the formulation proposed using the weights, skills, measures, and the three mathematical formulations.

d. The Numbered Fleets reviewed the measures, OE categories, and A, P, E functional areas for applicability and validity

e. Two different teams assigned the current set of measures to both OE categories and APE major functional areas, these assignments will be compared.

viii. Cost data categorization and JMET assignment validity

a. U.S. Fleet Forces Command, Commander Second Fleet, and JFCOM J-9 provided cost data for this research.

b. Each of the above reviewed the categorization of the cost data into the four categories proposed. Only Commander Second Fleet was able to provide data for all categories.

c. Each of the staffs which supplied data also reviewed the JMET assignments to the costs for validity.

ix. Validate models during operational staff events and certifications

a. The primary candidate for observation and validation was Commander Second Fleet, which participated in FA PANAMAX as a part of a CJTF certification during September 2007.

b. JFCOM J-9 provided an opportunity to use the performance assessment and costing methodology during Noble Resolve 07-2 conducted during august 2007.

c. A final opportunity to use the performance assessment methodology for a specific event was afforded by JFCOM J-9 during exercise TOP OFFICIAL (TOPOFF) 04.

x. Analyze results

a. Perform a mathematical analysis of methodology results

1. Validate that the measures, organizational elements, and major functional areas have agreement to within 10% of the initial proposed sets.

2. Two teams will develop the measures, organizational elements, and major functional areas. Additionally, the five numbered Fleet staffs will review each of these and submit their agreements or differences. The 10% value was chosen in order to provide a small enough data set (which would be almost 100 measures for that category) to correct within the period of this research.

3. Perform an analysis of the results of the values for the overall Organizational Element results obtained by the mathematical formulations proposed to determine the utility of each method. Results should be within 5% to be considered useful.

4. Perform an analysis of the results of the overall Organizational Element results when linked to the APE functions to verify that the ability to link these values is acceptable for the either weighting approach. The values should agree within 5% to be useful.

5. Perform an analysis of the values obtained for both the LDF method for either the Organizational Element values or for the APE functional areas as compared to either weighting approach. The values should agree within 5% to be useful.

6. Perform an analysis of the Second Level Aggregation as a means of linking Organizational Element results to APE functional area results. These values should agree within 5% to be useful.

7. Verify that performance curves for successive normalizing weights retain the same shape and relative change as a result of the new weighting values.

b. Perform Subject Matter Expert (SME) analysis of methodology results

   1. Since the application of cost data to specific UJTL categories is not implemented, application of historical exercise data, where applicable to the methodology developed was be reviewed by SMEs (such as the CFFC staff) to determine the validity of the approach.

2. Similarly, there is no formal method of determining ROI, the formulations of ROI developed herein were used with available historical data and compared to any previous ROI calculations to validate the utility of the methodology.

3. The output of the performance assessment methodology compared to the verbal assessment of the FA PANAMAX provided by the JFCOM Observer Trainer team. These results were presented to Vice Admiral Chanik for comparison to the numerical values obtained through observation by the assessors assigned by the Second Fleet staff during FA PANAMAX 07.

4. The output of the performance assessment methodology as it applied to Noble Resolve 07-2 and TOPOFF 04 was reviewed by the director of the Joint Context division from the J-9 Directorate of JFCOM.

3.3.2 Experimental Design and analysis of results

This section describes the experimental process that was utilized to evaluate the effectiveness of the performance assessment, costing, and ROI methodologies developed in the course of the research. As has been discussed in previous sections, the validity of the measures chosen, the appropriateness of the skill areas, the validity of the OE categories, the comparison of the three mathematical constructs for evaluating performance assessment, and the accuracy of the assessment and cost methodologies for formulating ROI require testing and analysis.

3.3.2.1 Measure, skill, and category validity

There are three fundamental issues regarding measures, skills, and categorization that must be addressed regardless of the ultimate algorithm and methodology adopted. The first area that must be addressed is the fact that objective measures are just beginning to be accepted as a part of performance evaluation, and as such are also just beginning to be developed for training and readiness reporting and use in certification. As such, there

are either developmental sets or sets that have been used in limited scenarios. All such measures were gathered and reviewed. Earlier in this dissertation there was a discussion regarding a number of measures that have been developed in parallel. For those measures that differed significantly (25% has was selected as the value that would be significant based upon the researcher's operational experience) resolution by means of agreement between originators was accomplished to permit the use of a single set of standards.

The validity of the measures will be tested in three ways:

1. By review by the originating organizations;
2. By review with operational staffs, primarily the numbered Fleet staffs;
3. By testing the measures in simulations to note impacts on overall OE or staff performance.

The validity of the OE categories and APE major functional areas were confirmed in several ways:

1. By comparison between the two groups initially assigning measures to either an OE category or an APE major functional area;
2. By review with operational staffs, primarily the numbered Fleet staffs during an initial assessment of certification capabilities;
3. The OE categories and APE functional areas were reviewed by personnel from the Synthetic Training directorate at Tactical Training Group Atlantic;
4. By review with Subject Matter Experts (SMEs) assigned to assess the performance of the Operational Level staff during FA PANAMAX 07.

All areas requiring either correction or change were noted and the results discussed in the dissertation;

5. By review with the SMEs assigned within the JFCOM J-9 directorate assigned to conduct Noble Resolve 07-2;

6. By review by personnel on the Territory of Guam who conducted both Noble Resolve 07-2 and TOPOFF 04.

The validity of the skill areas used to bin measures for a given OE category or APE major function will be confirmed by:

1. Applicability of the skill area to all three possible approaches,
2. Review with operational staffs,
3. Review with SMEs from Commander Second Fleet responsible for devising the applicable measures for FA PANAMAX 07,
4. Review of the SMEs from TTGL who were involved in the execution of Fleet Synthetic Training (Joint) 07-2,
5. Review of the SMEs from JFCOM J-9 directorate responsible for execution of Noble Resolve 07-2 as well as observing TOPOFF 04.

3.3.2.2 Comparison of Performance Algorithms against each other

The first comparison of the algorithms for performance assessment was conducted by reviewing the results each algorithm produced against the other two in a series of simulations. The results in this testing were used to verify the choices for skills, the weighting of each "bin" or OE category when used, and the normalization factors when applied. The results for the logarithmic functions followed the characteristics delineated in the methodology section with regard to performance improvement or decline,

insensitivity to single measure variations, and a change in slope that represents human learning as performance improves.

The differences between single aggregation and two-step aggregation were noted and the most appropriate aggregation method will be retained for use in the operational staff evaluation. The method of determination of "goodness" was the ability to represent change, in successive days when possible, in a consistent manner as reviewed by operational staff subject matter experts. The staffs were presented with results, when possible, for multiple observations representing a progression of training. The staffs were also presented with the results of trials as might be expected during an exercise scenario so that they were aware of the impacts of the variations of single measures, declining or static performance on overall staff performance, and overall performance trends.

A secondary feature of this comparison was that the various staffs were able to view the impacts of weighting factors assigned based upon the specific scenarios presented. These scenarios were significantly different, such as the importance of liaison efforts in scenarios where major combat operations were subordinate.

The impact of choices for T-4 values, which varied from 0 to as high as 65% as determined by the SMEs were also available for analysis. The numbers of measures available to assess performance in the APE categories was also investigated during this research and provided to all the staffs involved.

3.3.3 Comparison of Performance Algorithm Results with Actual Staff Observations

The final comparison to be made was that of the performance algorithm's output of performance assessment during the course of an actual exercise or certification event against the actual performance of an operational staff. The data from the previous discussion of methodology comparison should minimize the differences between the performance evaluations the staff performs and those generated by the methodology. This step may be undertaken on multiple occasions until the difference between the methodology and the staff's self-evaluation is within a reasonable margin, which will be initially forecast as no more than 10%. This value has been chosen because there are

factors in team interaction, staff integration, and even staff rotation that preclude a measurement that might take place within a 3 month period from being identical, even in a repeated scenario. Early discussions with COCOM and Service Headquarters staffs indicated that even an 80% agreement in methods would be far superior to anything currently in use. There are currently very few (the exception being PACOM) exercises that are objectively assessed.

3.3.3.1 Validation of Cost Algorithm Methodology

There are currently no formal methodologies that formally relate costs incurred in either live or large-scale distributed synthetic events. The validation of the algorithms developed will be against historical data that is by review by SMEs from at least two COCOM or Service Headquarters staffs for applicability to the selected measure, JMET, OE category, APE major function, or lastly event.

As with the comparison of performance assessment methodology during major exercises described above, cost data was obtained from three different exercises and compared to (in the case of data for Major Combat Operations) against historical data and historical constructs (fitting cost data to the cost modules and categories) to validate the algorithm and output for long-term use as a formal method.

3.3.3.2 Validation of the ROI formulations

There are currently no accepted or routinely applied ROI formulations that are in use. The formulations described as the initial possibilities as well as others that might be requested during the completion of this research will be presented to various COCOM, Service Headquarters, and Operational level staffs for applicability, acceptability, and utility. Though various formulations may be mathematically accurate, the ultimate utility of any of the formulations devised will be the potential for use in defending resource allocation or resource requirements. A secondary measure of acceptability and utility will be the ability of the organizations listed to construct more uniform and more formal resource requests or defenses which can be applied to all the organizations listed without need of Service specific terminology.

3.4 **Research Limits**

The goal of this research is to produce reliable and repeatable methods for portraying ROI in a any training or operational environment. The current implementation of certification processes requiring more quantitative measures, the lack of formal costing methods that specifically address training readiness, and the need to more uniformly request resources for training and readiness, notably investments in synthetic training capabilities are all motivations for this research.

The decision to use the UJTL as the basis for both performance assessment and costing does present some possible controversy. There are always discussions regarding the definition of conditions and standards, and these discussions can dramatically change the nature of any assessment. This research uses the position of the Commander, in other words, the operational title as the condition. This limits the possible missions for which the JTF can be assigned, and is consistent with the Combatant Commander requirements for operational level staffs.

The standards required to complete the assigned mission can now be specified in a very specific context and quantified to permit assessment. There is no intention to assert that a change in condition will occur without a change in the applicable and relevant standards, indeed the standards must be reviewed for each change in condition. It is even possible that standards may require adjustment for a given condition should changes in specific METs be required by a higher headquarters.

The choice of this framework for conditions and standards permits any staff or organization in question to focus on the METs and supporting sub-tasks, assessing the importance of each, the skill sets required to achieve the tasks, and ensuring that the attributes used to quantify achievement are correct.

There is also the natural tension between arriving at a single value for performance that appears to yield a value for performance, as might be expected for a physical phenomenon such as voltage, current, power, or work in engineering analyses. This research will attempt to characterize performance in numerical quantities, but until there

is consistent use of either this methodology or one that is similar, so that attributes, weighting factors, and skill assignments are well understood, the values which are used to portray either Organizational Element or Functional area performance merely the initial data points providing insight into the performance trends of operational level staffs.

The methodologies developed in the course of this research, as well as the insights gained in performance assessment, costing linkages, and formulations of ROI are certainly not an end-state by any means. Rather, the results of this research should provide a platform for more accurate assessment measures, the refinement of existing measures and standards to address areas within the JMETs that are currently insufficient in detail to portray operational staff performance, and perhaps insights into staff organization, process or training. Similarly, this initial attempt to relate costs to specific training and readiness standards should provide a template for further cost association with training and readiness so that any event can directly show the training increases due to resource allocation, or projections for costs to prepare a staff for a specific mission can be more accurately forecast. Finally, the formulations for ROI are intended to be a template for expanded use to validate the power of synthetic training as well as the expenditures to date on large-scale distributed simulation.

The development of this methodology in an unclassified domain is specifically intended to facilitate use and dialogue between operational staffs undergoing certification and between COCOMs or Service Headquarters to standardize methods to report and portray readiness and costs. The migrations to a classified domain so that direct application to other available tools that permit the identification of training or manpower deficiencies and the requisite training to improve performance against specific classified Operational or Contingency plans is one possible offshoot of this research.

A secondary, but no less important reason for developing this methodology in an unclassified domain is the possible use outside of DoD. Throughout the dissertation, various applications to industry, government, or even non-government agencies have been suggested. The authors cited have noted that the same lack of ability to quantify performance for training, to accurately portray costs associated with training, and to

present ROI values that are compelling to either corporate or government leadership. The very simple examples of state or municipal disaster preparedness training, training to combat or recover from a possible terrorist action, or a large corporation's need to coordinate geographically dispersed actions are all variations of the scenarios used by the U.S. military. The concept of METs is applicable, with some variation, as is evidenced by the introduction of the Department of Homeland Security (DHS) Universal Task List (UTL) (64). State Emergency Management Agencies (EMAs) function nearly identically to operational level staffs in that there are core functions, organizational elements, and METs that must be accomplished to overcome a specific type of emergency. The conditions and standards discussion above are applicable to this situation in every regard. Further, the "intangible aspects" of decision-making are equally applicable in or out of a military context.

Industry also has contexts which could utilize the conditions and standards discussion. Complex simulations, such as those used to train pilots, can be characterized in terms of the experience of the pilot, thus stipulating a condition. The standards for a variety of scenarios may (but are not necessarily always) dependent upon that experience level. So, once again, the attributes, in this instance, some known quantities such as power, rudder changes, or aileron position can be stipulated for satisfactory performance. The weighting factors and skill sets can be identified for either all, some, or only specific experience levels. Variations on the methodology proposed here could provide similar benefits to all the organizations cited as are projected for the U.S. military.

Regardless of the methods eventually developed and used, the requirement to more clearly portray the appropriate ROI parameters that support training U.S. forces, specifically CJTF or FCC staffs, in the most robust possible training scenarios possible is the goal. A secondary goal is to show that in the vast majority of cases distributed synthetic training is not only sufficient, but it should be the preferred method to prepare such staffs due to the much greater range of scenarios, more robust opposition force compositions and capabilities, and greater ease of implementation. A final goal is to demonstrate the fact that any SJFHQ staff requires nearly identical assessment criteria

that are not Service dependent, since their ultimate mission is to coordinate or command (depending upon whether the role is ultimately a CJTF or FCC staff) forces that are typically assembled from across the Services or from Allied nations.

None of these goals is achievable unless culture permits honest evaluation and formulation. In the end, the Services do not compete with one another, though budgets are allocated that way. The ROI realized from either live or synthetic training has only one object, to enable the employment of U.S. forces quickly and effectively, whether the need is humanitarian support or combat operations. We have to know our enemies, but first we MUST know ourselves, then we can fulfill the final part of Sun Tzu's immortal words regarding the art of war, to fight and win 100 battles (65).

4 MODEL CONSTRUCTION AND IMPLEMENTATION

4.1 Performance Assessment Formulation

The development of the ultimately adopted Performance Assessment tool required several iterations, which, owing to opportunities that were unexpected and driven by deployment or already existing exercise schedules often occurred in parallel. The development of the final model was completed based on lessons learned from two previous attempts and in concert with a certification effort for an existing operational level commander.

4.2 Operational Level Command Certification Model

4.2.1 Opportunity Presented for Certification Event Application

Formal certification of an Operational Level Commander as a CJTF within a specified range of missions as delineated by that Staff's Combatant Commander (COCOM) is just being implemented by JFCOM. The first Operational Level Commander to undertake the certification process was well known to the researcher and the Commander agreed to allow the researcher to assist in the certification effort as well to utilize any assessment (as well as the cost data necessary to complete the event) tools devised as part of the Command's overall assessment of performance.

The process of identifying the JMETs as well as appropriate measures and initial performance assessment tools began in April 2007. The exercise took place between 31 August 2007 and 6 September 2007. Two previous events had provided some insights in the formulation of the model used for the Operational Staff, and those insights will be detailed later in this section. Those events had assisted in the identification of the various OE and APE functions that were common to almost any operational level staff, and this staff was one of nine staffs of similar type that had previously agreed with the assignments from the UJTL as well as to the standards that had already been developed. The certification effort was a greater challenge than had been anticipated in the April time-frame, but those challenges ultimately proved useful to this research.

4.2.2 Initial Model Theory: Use Existing JMETS

The initial formulation of the performance assessment tool was based upon documentation supplied by the involved staff. This listing only portrayed the two digit JMETs that would be required to complete the certification effort during the course of the exercise. A review of this list of JMETs resulted in a proposal for some additional 2 and 3 digit JMETs to be added to the list which would support the primary JMETs for certification. These were retrieved from the RATIONAL REQUISITEPRO data base and then set up in the performance assessment tool format by either OE or APE as appropriate.

Since this would be a four day exercise, and would be evaluated on each day, the need to have consistency for day to day observations and to see trends made it necessary to ensure that changes in performance, even if they were projected trends, would be consistent in all four days of evaluation. Owing to the problems encountered and discussed in the formulation of the Major Combat Operations performance assessment tool, it was also imperative that there were no redundant measures from one OE category or APE category to another. The result was a performance assessment tool contained 813 measures. The proposed team of assessors that would be provided by the Operational Level staff in question was to be comprised of 25 members. Using the guidelines discussed in later in this dissertation, an assessment team of this size e required each observer to be responsible for a little more than 30 observations each, which could have proved to be more than could have been reasonably accomplished by an inexperienced assessment team.

The proposed assessment tool was submitted for review and modification in mid-June 2007. The time between submission and the exercise was occupied with the two events, a Major Combat Operations (MCO) performance assessment and costing effort and an experiment which was concerned with Defense Support for Civil Authority (DSCA) and Humanitarian Assistance/Disaster Relief (HA/DR), as well as testing the response of the certification exercise assessment tool. The results of that testing will be discussed in a subsequent section of this dissertation.

4.2.3 Change due to Training Objective Approach

Communication with the staff through the months of July and August gave no indication of the lack of acceptance of the performance tool submitted or that there had been a change in the scenario and JMET requirements to be addressed during the certification. Both of these issues were revealed early in the week preceding the exercise. The staff had reviewed the proposed performance assessment tool and had decided that the training objectives requiring completion to support the JMETs specified by the COCOM would not be adequately addressed with the proposed performance assessment tool. The JMETs specified by the COCOM had expanded slightly from the listing originally discussed, including some strike, small scale contingency operations, and some expanded Foreign Humanitarian Assistance missions.

In order to address the expanded requirements from the COCOM, the Operational Level staff met on several occasions and developed a list of training objectives that centered on each of the specified two digit JMETs

This listing was based on two documents, a training template devised by the Operational Commander and the COCOM approved Joint Training Plan (66). An excerpt from the COCOM approved Joint Training Plan is available in Appendix D. An example of one of the training objective descriptions is included in Figure 11.

Using these two documents, the staff created a collection management plan specified the assessor, key JMETs, and levels of performance required. The levels of performance specified in both the training template and collection management plan were the basis for developing the training objectives. As is the case with the UJTL, there was a combination of both objective (quantitative) and subjective (assessed as a binary measure, Yes/No) measures.

Using these two documents, the staff created a collection management plan that specified the assessor, key JMETs, and levels of performance required. The levels of performance specified in both the training template and collection management plan were the basis for developing the training objectives. As is the case with the UJTL, there was

a combination of both objective (quantitative) and subjective (assessed as a binary measure, Yes/No) measures.

TO Number 4006 Fuerzas Aliadas / PANAMAX 07

Supporting Tasks: OP4, OP 4.4.2, OP4.4.2.2

TO Performance: Develop, distribute, and monitor compliance with policies and procedures for personnel support and reporting.

Training Audience: OSI, JPC, JRC

Training ***(1) Method*** -Academic Training ***Mode*** - Lecture ***Media*** - Slides

(2) Method -Exercise Training ***Mode*** - CPX ***Media*** - CAX/MSEL

Focus Area: Personnel ***Mission:*** Stability & Support Operations

Responsible Org: OSI ***JCLL Title:*** Core, 6150, 6175, 6225, 6333, 6400, 6601

Training Situation:

1. JTF established.
2. Orders received.

Level of Performance:

1. 85% of required component personnel services are in place and operational.
2. Replacement and rotation personnel obtained and assigned with minimum impact on operations.
3. Personnel estimates are part of all plans development and assessment.

Figure 11: Sample Training Objective

The training objectives were cited on several spreadsheets and were numbered to correspond with the overall JMET that was being supported which was based upon the same numbering scheme used to formulate the training objectives, e.g. 4006 above specified training objective 6 for JMET 4.0, Provide Operational Logistics and Personnel Support (1). The listing the staff ultimately compiled consisted of 404 training objectives

covering the JMETs considered necessary by the COCOM for certification as a CJTF. The "core" JMETs are listed below:

OP 1.1 Conduct Operational Movement

OP 1.5 Control Operationally Significant Areas

OP 3.1 Conduct Joint Force Targeting

OP 4 Provide Operational Logistics and Personnel Support

OP 5.2 Assess Operational Situation

OP 5.5 Establish, Organize, and Operate a Joint Force Headquarters

OP 5.7 Coordinate and Integrate Joint/Multinational and Interagency Support

OP 6.2 Provide Protection for Operational Forces, Means, and Noncombatants

OP 6.5 Provide Security for Operational Forces and Means (1)

Each of these had multiple supporting JMETs as shown in Figure 11, and there were specific training objectives, shown in Figure 11 as well, or measures of performance that were devised by the staff to assess achievement of each training objective. A full listing of the training objectives is not included in this report, but an excerpt from the spread sheet devised by the staff to: 1) Number each for later recording purposes, 2) Identify key operational sections of the staff, and 3) Provide the verbiage for each training objective is shown in Table 5 (The reference for all the terms in the accompanying footnote is Joint Publication 1-02 (67)).

JOC	**Monitor the status of assigned forces and resources and execution of directed actions**	
11021	Maintain COP with minimal track conflict and data latency	JDN
11022	Maintain SA within the JOA: friendly, MN partner, IA, NGO/IO, enemy	JDN
11023	Sustain collaborative architecture that allows for a quick and timely flow of information between the JTF operating forces and headquarters	JDN
11024	Provide appropriate managed data feed to friendly partners as required	JDN
11025	Monitor and track the execution of directed orders	JOC
11026	Manage operational RFIs	FOPS

Table 5: Sample Training Objectives and Descriptions [18]

The complete set of Training Objectives and Descriptions was presented to the researcher on the afternoon of 29 August 2007, three days before the exercise was scheduled to start. The major concern with the receipt of the entire document was that, as is illustrated in the example above, the Training Objectives were specified. In many cases the numbering scheme was completed, and again, in many cases the cell or organization responsible for the action was identified, but there were no assessment standards included for any of the Training Objectives.

4.2.4 Development of the "Rosetta Stone" and Observer Sheets

The very compressed time between Training Objective delivery and the commencement of the exercise presented a significant challenge. Each organization responsible for assessment had provided input as to what would be assessed and the description of the task, but with only a few exceptions, there were very few objectives

---

[18] JOC – Joint Operations Center, JD—Joint Data Network, FOPS- Future Operations, SA – Situational Awareness, JOA – Joint Operations Area, MN – Multi-National, IA – Interagency, NGO/IO- Non-Governmental Organization/Information Operations, RFI – Requests for Information

embedded within the descriptions, to use as a basis for assessment. Further, the personnel who were assigned to perform assessments were generally not familiar with how to conduct assessments from previous duty assignments, so a format was required that provided a single point of reference for all the inputs from each assessor was desired. The format needed to include, at a minimum: 1) a very clear description of the performance standard (the "passing grade"), the range of numerical values that would portray better than average (T1), below average (T3), or unsatisfactory (T4), 2) the JMET, 3) the associated measure from the UJTL, and 4) a set of columns which could be used for another reporting requirement, inputs for Navy Lessons Learned program.

The one difficulty that had been encountered in previous formulation attempts, especially in the Major Combat Operations model, redundant measures, was precluded by the numbering system devised by the staff. The issues that required resolution in a bit over 72 hours were: 1) Devising the performance standards and range of numerical values that would apply (or assignment of binary criteria), 2) Completion of the numbering scheme for all measures, 3) Completion of the assessor assignments, and 4) Creation of the performance assessment (and by extension the cost tool) tool.

The key to all of the above was each Training Objective's description. As noted in section 4.2.3, with only a few exceptions there were very few direct numerical values that resided within the Training Objective descriptions. There were some of the descriptions that were very obviously subjective, but the remaining Training Objectives would require complete assignment. Additionally, although a standard may have been provided, in terms of percentage or time, only the standard to be achieved was available, but not any range above or below that standard to provide a means of describing what was within acceptable limits for achieving the standard (e.g. +/- 10% of the specified standard) or what would be well above, below, or unsatisfactory in comparison to the standard.

In order to implement the performance tool that had been developed for the other three situations, each Training Objective would require weighting, OE and APE assignment, and skill assignment. These would all require agreement from the staff, and,

with very little time available to complete all the necessary steps required to have finished products available to the assessors and be in a position to collect data for performance assessment some agreements with the staff were reached:

- The weighting for each Training Objective would be executed in a manner consistent with the framework described in a subsequent section of this chapter, H/M/L. It would be done within each Training Objective grouping, as represented in the JOC example of Table 5.

- The JMET and measures would be assigned such that if the wording in the Training Objective description was either exact or very similar, the JMET, measure, and standards would be used.

- Since the responsible board, center or cell had been assigned in almost every instance, these would be translated into their major organizations, so that the standard organizations from Joint Publication 3.0 would be the basis for Organizational Assessment

- The descriptions would be used to assign the most appropriate APE function to each Training Objective

- The LDF weights would be assigned as devised in a subsequent section of this chapter, using the most operationally significant skill assignments as the basis

- The best way to handle the subjective measures was to assess them as either at standard, or at some value that would be unsatisfactory, but non-zero

- The personnel who should ultimately make the judgment as to the correctness of the proposed standards (both objective and subjective) should be the assessors, who could request changes during exercise execution.

All of the above were agreed to, and a format, referred to as the "Rosetta Stone" was devised almost immediately. All of the elements described above were included. In order to ascertain whether the construction of the "Rosetta Stone" as well as the Observer Grade sheets would be satisfactory to the staff, a sample for the OP 4.0 JMET was supplied to the staff. The format, the assigned weightings, JMET assignments, OE, APE, and skill assignments were also accepted. This did not constitute final acceptance, but an agreement that the methodology was logical and appeared workable. Further, the prototype observer sheets were also accepted. Examples of the "Rosetta Stone" and Observer Grade sheet are provided in Appendices A and B. The early completion of these two documents also permitted briefing to the Senior CJTF staff as well as the Commander so that they were aware of the formats and approaches involved.

4.2.5 Translation to Performance Assessment Tool

Completion of the "Rosetta Stone" and the Observer Grade sheets was the priority until Saturday, 1 September, 2007. The assessors needed to have the formats in hand and receive training prior to beginning assessment on Sunday, 2 September 2007. The very compressed timeframe did not permit extensive matching of the Training Objectives to JMETs as would ultimately be required for exercise reporting requirements, but if the specific Training Objective was easily matched, then the assignment was made, if not, then the assignment was completed either during the course of the exercise or after exercise completion.[19] Regardless of whether an applicable JMET could be easily identified and the accompanying standards assigned, all the Training Objectives required standards. One further complication to the process was also discovered during the effort to devise or assign standards to each Training Objective, which was the relevance of the standards which did exist for some of the relevant JMETs. The issue was evident in many of the time or percentage based standards. Though the numbers didn't appear inordinate in isolation, many of the values seemed to be more applicable to a more

[19] The exercise and certification reports were not due until 1 October 2007, nearly four weeks after exercise completion.

tactical application. The values which related to either liaison or disaster relief required modification based upon recent lessons learned from operations such as the recovery from Hurricane Katrina.

Similar issues but in much smaller numbers had been raised during discussions with the nine operational level staffs during initial performance tool construction. There were concerns that some of the time standards didn't appear to account for the time constraints imposed on staffs as a result of higher headquarters reporting and coordinating requirements, the time constraints of other component commander staffs, or the very simple time requirements that existed due to the CJTF's own daily requirements. All of these factors were taken into account when either devising new standards or when assigning them from existing standards.

Once the standards and grade sheets had been completed as required, it was necessary to construct the performance assessment tool. The creation of the "Rosetta Stone" made this task immeasurably easier than in any previous instances.

Since the OE, APE, weighting, and skill assignments were already included on the "Rosetta Stone," converting this format into the one required for performance assessment was simply a matter of using the tools available within EXCEL™ to copy and paste the appropriate JMETs into the corresponding OE and APE categories. The staff's numbering system also made sorting much easier than in previous work, so the only real time spent in the construction was setting up the various calculations for each area. This portion of the preparation for the exercise required less than eight hours to complete and test. The entire effort from commencement to being ready to enter performance data required approximately 30 hours. The task of data entry was simplified greatly because each observer also had an EXCEL™ version of the Observer Grade sheets and these could be emailed directly to both the staff exercise coordinator and to the researcher.

4.3 UJTL Usage for the Performance Assessment Model

The use of the UJTL to identify the missions and supporting Joint Mission Essential Tasks (JMETs) is well understood by Commanders at the Strategic, Operational and Tactical levels of operations. The actual use of JMET standards as an assessment tool for performance whether that performance is in the course of a certification (to ascertain the proficiency of the staff in question), or as a guideline for training is not well established. The requirement to report via the Defense Readiness Reporting System is established and being followed, but until October 2007, use of JMETs to report in the system has been done primarily in a traditional "stoplight" fashion.[20] In December 2006, a local Operational Commander expressed a desire to tie doctrine, the UJTL and other Joint Publications to processes, and finally, to numerical standards. The Commander agreed to utilize the methodology being developed by the researcher in conjunction with the first Quadrennial Defense Review directed Joint Task Force "Capable" Headquarters observed certification event in September 2007.

Based upon this discussion, an investigation of existing JMETs was undertaken. Three listings had been identified between September 2006 and the December 2006 meeting. These had been developed in support of the Operational Level Commander to be certified (but not specifically due to the certification effort), to support the Blue Flag exercise series conducted by the U.S. Air Force which conducted training events for Numbered Air Force Commanders in preparation for deployment where the Commander would be executing operations as a Joint Force Air Component Commander (JFACC) or in support of the development of the Concept of Operations (CONOPS) for the Standing Joint Force Headquarters effort. These efforts and documents have been discussed in various parts of Chapter Three of this dissertation.

---

[20] Stoplight charts refers to the red, yellow, green terminology depicting unready or incapable, ready but degraded, and ready. The ability to show variations to each of these is problematic, but done with appropriate verbiage.

A database was created using RATIONAL REQUISITEPRO™ offered to the researcher by Joint Forces Command J-7 which listed every JMET from the UJTL. Once this had been done, the standards that had been identified from the three sources available were added where appropriate. As was also discussed in Chapter Three, there were 105 instances where at least two of these three organizations had developed standards for the same JMET measure.[21] Of these measures, 53 used the same range or subjective value and 52 had some differences between them. Further investigation of the 52 with differences reduced the number that required some form of resolution to 30, which had differences in the attributes (percentages, time, or instances) that exceeded 25%.[22] In almost every one of these instances, the reason for the difference was primarily because the organizations developing the standard failed to either communicate with each other or believed that the standard developed would be more appropriate for the JMET achievement. These differences were resolved by contacting the organizations in question and proposing a value that could accommodate all parties.[23]

4.3.1 UJTL Breakdown by Category

While conducting the research required in developing the performance assessment tool, a related project came up with the Operational Level Command preparing for certification. The capability to utilize various minimally tasked staffs to support a

---

[21] There were a total of 972 individual measures identified between the three organizations specified out of the total listing of 2742

[22] The 25% value was chosen because most of the differences were less than 10 to 15% (a standard of .80 vice .90 was selected for example, there were only 30 instances where there were significant differences in assigned standards).

[23] During the final stages of this research, the SJFHQ CONOPS were published in final draft form. This draft deleted any reference to standards for specific JMETs, and the eventual development of the performance assessment tool for the Operational Level certification did not use any values which had been previously devised by SJFHQ, at their request.

headquarters that was heavily tasked in order to reduce the work load of the engaged staff was being investigated and developed. This concept known as the Maritime Headquarters with Maritime Operations Centers (MHQ w/MOC) would involve not only Numbered Fleet Headquarters, but several other Higher Headquarters including both four-star Navy headquarters. The concept involves devising a core set of JMETs which all MHQ w/MOC can perform and then create tailored sets of capabilities which relate to each Headquarters' unique operational responsibilities (67). A key aspect in devising these core capabilities was the investigation of each Headquarters' expected roles and missions. These were viewed across the Range of Military Operations (ROMO) in four categories: Normal and Routine Operations, Foreign Humanitarian Assistance/Disaster Relief, Stability Operations, and Major Combat Operations. In order to arrive at what the "core" capabilities each of the above consisted of, it was necessary to examine the UJTL for JMETs that would comprise the responsibilities for any organization in each of the four categories named as well as their responsibilities if they were to act in the role of a Service Component, a Joint Maritime Component Commander, or a Joint Task Force Commander. Two different groups examined a series of documents including the MHQ, w/MOC Concept of Operations (68), Joint Publication 3.0 (2), the SJFHQ Draft CONOPS (39), and the suggested Joint Mission Essential Task List (1) for both a Joint Maritime Component Commander (JFMCC) and CJTF. A total of 134 JMETs were submitted as the entire set required for any role selected across the ROMO discussed above. In order to understand how any organization would implement these, it was deemed necessary to separate each of these into the appropriate Organizational Element or Assess, Plan, or Execute (APE) major functional area category. Both groups would investigate the list and any possible supporting JMETs, which ended up including the all of the 972 JMETs which had developed standards. The process discussed above and the resulting formulation of the performance assessment tool were also discussed in section 3.2.2.1, however additional detail which was a natural consequence of the proposed methodology and which was discovered during the research will be expanded in subsequent sections of this dissertation.

4.3.1.1 Organizational Elements (OE)

Regardless of the specific missions that any of the headquarters would be assigned in any of the possible roles, and regardless of the role that any headquarters would be executing, each had a basic set of organizations that would have subordinate boards, centers, bureaus, and cells to execute the JMETs necessary. These organizations are identical to those discussed in Chapter Three and depicted in Figure 10. Both groups agreed that since these were based in doctrine the Organizational Elements should be the ones discussed in a series of visits to nine different Headquarters and also be assigned JMETs so that the volume of responsibility for any particular OE in any role or military operation category could be better understood. It also seemed apparent that since boards, centers, bureaus, and cells comprised individuals from all parts of any command, that there could very well be instances where a JMET could have multiple OE assignments. This was entirely acceptable since almost all participants had served on military staffs in varying capacities and understood what sorts of tasks were required within any OE subgroup.

4.3.1.2 Assessment, Planning, and Execution (APE)

The same logic that was applied to assign OE categories to each JMET was applied to the APE categories. There was concern that although the combined group working on assigning both OE and APE categories were comfortable with the fact that these assignments were necessary, there were many more Execute type JMETs with standards than either the Plan or Assess categories. The concern was how to minimize the importance of Execution at the operational level since; even if Execute was defined to include the Execution of orders and the Monitoring of Execution, this activity should comprise no more than 10 – 15% of an Operational Staff's time. The numbers of Execute type JMETs comprised 70% of the 134 JMETs under consideration as well as 55.6% of the 597 measures required to assess these JMETs. In contrast Assess, which is a primary activity of a staff, especially in reaction to any feedback from operations, comprised only 32.8% of the JMETs and 27.8% of the measures. The group was in general agreement that at the operational level, the staff should be spending approximately 60% of its time in planning, 25% of its time in assessment, and 15% of the

time in execution. There did not seem to be an effective way to limit the Execute type JMETs, but the fact that assessment had so few JMETs was somewhat surprising. The lower numbers of assessment JMETs to Plan and Execute JMETs did not seem to be an issue at this point.

The two groups reviewed each of the 597 measures and assigned OE and APE categories to each. The groups worked in teams of three which provided balance to the assignment process due to the collective distribution of operational experience in each group. When completed, 1995 assignments were made. The groups then agreed that each should review the other's assignments to ensure that nothing had been missed and that no bias from any single individual in either group would unnecessarily influence the assignments. After review, there were 110 assignments that were changed from the original assignments, which reflected a 94.5% agreement between the groups.

The assignment of OE and APE labels to the JMETs was critical to forming the performance assessment model since this task would have been very time consumingfor a single researcher to perform. It also provided a valid basis for dividing the JMETs into the proper categories.

The visits to the nine headquarters staffs did not change the assignments significantly. The visited staffs sometimes felt that a particular OE might not capture a specific JMET with the assigned measures, but there were no instances where the JMET was found inapplicable. A very similar trend was noted with the assigned standards. The team that visited Pacific units was advised of 10 instances where the measures were not appropriate for the Operational level, but the team visiting the Mediterranean and Middle East commands had no issues reported with the attributes or standards that had been applied to the measures by the work groups.

The results of the visits and assignment processes made construction of the initial performance assessment tool considerably easier and added confidence to the standards and organizational categories for the JMETs.

4.3.2 Binning Assignment Process

The issue of single skill assignment as opposed to multiple assignments was similar to the issue raised with the OE and APE assignments. In this case however, since many of the JMETs had specific characteristics such as accuracy, depicted in the example in section 3.2.2.1, there did not appear to be any compelling reason to assign more than one skill per measure. The descriptors for the skill assignments in the initial formulation included 28 possible choices. The definitions of each term are supplied in Appendix C. The candidate skill characteristics are shown in Table 6, which expanded the listing discussed in section 3.2.2.1.

These possible choices appear to cover a broader range of possible assignments than the original listing in section 3.2.2.1, which could also better describe the range of desired characteristics which could be exhibited by an operational level staff and Commander.

Acceptability	Accountability	Accuracy	Adaptability	Agility
Authority	Balance	Clear Communications	Confidence	Competency
Courage	Decisiveness	Efficiency	Flexibility	Initiative
Leadership	Liaison	Negotiation	Objectivity	Perceptiveness
Preparedness	Projection	Responsibility	Situational Awareness	Suitability
Thoroughness		Timeliness		Trust

Table 5: Skill Assignment Categories

4.3.3 Weighting Choices

The objective of using variable weights for both the Floating Weight Average and for the LDF method was that these would provide the Commander and staff the ability to decide what measures would be most important and assign values that would be reflective of that decision and in retrospect, was a significant element of model construction and implementation.

The Floating Weight values used for the operational staff exercise performance tool were based upon a High/Medium/Low scale. Values of .8/.5/.2 were assigned respectively.

The basic methodology, which was validated in the construction of the operational level staff evaluation tool for a certification exercise event, was to view high, medium, or low as follows:

- High – An essential task or function that must be completed to assure mission success. For each different situation, these tasks could vary. An example of that variance is the establishment of liaison with either a Non-Governmental Agency or international agency is extremely important during a Foreign Humanitarian Assistance situation, but it is not a significant issue while conducting Major Combat Operations.

- Medium – A task that must be carried out as a matter of routine to ensure information flow or the processing of information for daily guidance messages and briefs. There are typically multiple information paths and staff processes which provide redundancy, although these functions must be completed, short lapses will not impact operations.

- Low – A task or function which requires completion but is done as a situation requires or might be done on a periodic vice daily basis.

4.3.4 LDF Choices

The initial assessment tool LDF weights which ranged anywhere from a minimum value of 0.4 to a maximum of 0.9. These values were used recognizing that the product of several factors would determine the LDF score. The calculations for any given method, whether based upon normalized or floating weight averages could easily be carried out, but the range of values tended to lead to many measures being weighted at the higher ranges (nearly 73% at either .8 or .9) which, although may be desired by leadership, in almost any setting does not accurately reflect the realities of tasking priorities even in a very demanding scenario.

A consultation with Dr. Will Startin (69) revealed that when he had encountered a similar issue in his earlier research (23), he had resorted to a far simpler and more direct approach to weighting assignment. He used a method that assigned either a high (.8), medium (.5), or low (.2) weight to tasks. The work conducted by Dr. Startin was at the tactical level and this breakout worked fairly well in describing the importance of warfighting tasks, as well as some staffing tasks.

Adapting this methodology to the operational level appeared to have potential, and a method to perform such an assignment was encountered during the formulation of the performance assessment tools for the two events preceding the Operational Level staff exercise.

The values (0.4 to 0.9) that were assigned to the LDF factors described in the initial formulation did not seem to correlate well with the values for performance that were calculated by either averaging method, or even the limited averaging methods. Again, as noted in the preceding paragraphs, Dr. Startin suggested carrying over the H/M/L approach to the LDF factors was the most elegant solution.

This approach had merits when considering weighting factors for each skill area because the areas that were most significant in each OE or APE construct would be those that should influence the overall performance outcome. The decision of what areas should be considered important had some similarity to the assignments done in the skill areas. Those skill areas containing the most factors were those that typically comprised

the key characteristics of a specific area. Stated another way, though accuracy is important to those who are in the Current Operations (COPS) organizational element, these people are primarily watchstanders who maintain displays, communications links, and monitor events. Accuracy within the displays or communications is important, but at this level of command, it is far more important to have a current picture which is termed "situational awareness." Additionally, competency in establishing and maintaining the various displays and communications links is more important than the absolute accuracy of any particular data packet within either. It was not surprising to find that although almost all of the 20 skill assignment areas were represented in almost every OE or APE category, certain skill assignments dominated each area. Those with the greatest number of measures, especially if far and away numerically superior to the other areas (e.g. 25 to 30 or more measures where no other area had more than 10 – 15) were assigned a high value. In some cases, where the numbers of measures were not greatly different, there could be multiple high assignments.

The assignment of either medium or low values followed logically from the above, with a typical break point of about a 50% representation being medium and those with no more than 25% being low. This approach to LDF assignment was consistent and did ensure that either a very important factor or a combination of medium factors could significantly affect any performance outcome as had been postulated by Startin (23) while preserving the fact that many factors could be recognized. This approach does not preclude the prerogative of the Staff or Commander being evaluated choosing to select a skill assignment area with very few or even some lesser number of measures than an area already designated high as being important, but permits an evaluation based upon importance to a specific mission to be placed on any set of skill assignment measures. The significance of such flexibility in the assignment of assessment criteria cannot be overstated given the current tactical and operational training assessment/grading climate.

As the project progressed, this method very often resulted in calculations that were very close (sometimes as close as the 3rd or 4th decimal place) to the overall values. The other application of this method was to use only the LDF values from the limited set to form the LDF product. Again there were numerous instances where this limited set of

products yielded a result that was very close to either the two complete averaging methods or in some cases all four calculations. More discussion of the implications of this abbreviated form of calculation will be presented in the Analysis chapter of the dissertation.

4.3.5 Use of Analytical Hierarchy Processing (AHP) to link OE to APE Results

The issue of linking the OE to APE results as well as finding the most accurate method for determining the LDF weights led to use of a variation of the AHP method devised by Forman (40). The variation used for the final performance assessment tool was n a derivative of the "survey" method discussed in the literature. Though there had been no formal inputs to any of the JMETs used in this instantiation of the performance assessment tool, one thing was apparent. The relative importance of one area to another appeared to be correlated to the number of factors in each group. For example, if COPS had 211 JMETs in the Execute functional area, and Future Plans had only 25, could that ratio be used to properly represent the relative importance of each area to the Execute functional area? The two different approaches and the results are shown in Tables 6 and 7:

	Original (Survey)	JMET Driven (Total Measures)
COPS	.31481	.29104
FP	.03647	.06627

Table 6: AHP Factor Variations

The calculations which linked OE to the APE functional areas were carried out using both factors, primarily to determine if there was an advantage of using either method to determine the most accurate representation of the contributions of the OE assessment scores to the APE areas. The table below depicts the results of both methods to estimate the relative value of the OE values to the values calculated for the APE functional areas.

All values shown are the average of the normalized average and floating weight averaging methods

	Original (Survey)	JMET Driven (Total Measures)	APE Calculated Value
Assess	0.8341	0.8324	0.8296
Plan	0.8324	0.8305	0.8265
Execute	0.8320	0.8325	0.8239

Table 7: AHP Factor Results for Estimating APE Performance

The results appeared to indicate that the JMET driven approach could be utilized without any substantial penalty. This was attractive because the survey method necessarily requires a significant number of inputs from personnel which is typically both intrusive and does not necessarily reflect anything other than the individual's view of the area for which they have primary responsibility. The JMET driven approach eliminated the need for surveys and appeared to provide valid results. All subsequent model OE values were linked to the APE values using both methods, but the more the researcher used the JMET driven approach, the more appealing it became due to the diversity of the training audiences encountered and each audience's diverse views of the OE areas relative contribution to APE.

4.4 Lessons from Previous Model Formulations

4.4.1 Initial Performance Assessment Model

All of the issues discussed above were encountered in large part in the formulation of the initial model. This was termed the "Initial Bins" model only because it was the first attempt to use the skill assignment bins LDF method to calculate performance. The variations in LDF factors, implementation of the Limited Normalized Average, Limited Floating Weight Average (termed Limited Alternate Average in the

spreadsheets), and initial use of the Analytical Hierarchy Processing (40) approach to link OE values to the APE results were completed during this model construction.

The initial attempt to utilize the AHP method was to use the AHP values to assign LDF factors (this was prior to the H/M/L discussion with Dr. Startin had transpired). It was apparent, even using only the researcher's experience as the guidance for choosing the relative importance of one area (COPS) in contrast to another (FOPS). The 1,3,5,7 and 9 or their reciprocal values were used in this approach and the relative importance of each OE area was more easily seen in the resulting calculations. The LDF factors were assigned in a range beginning with 0.9 for the OE with the highest average, and then decreasing in 0.05 increments until all the areas had been assigned factors. This approach to assigning LDF weights was not particularly insightful because there seemed to be only a minimal distinction between the most important skill areas and the least. The lack of differentiation was also noted by Dr. Startin, so the High/Low/Medium approach was adopted for LDF factors. This was later modified to use AHP factors which were JMET driven as described in section 4.3.5

The Initial Bins model had several significant flaws, especially due to the redundancy of JMETs that undoubtedly caused the calculations to be somewhat suspect due to the inability to track (in all cases) the observed values for each JMET from one section to the next. A substantial amount of time was spent ensuring that these occurrences were minimized, and in the final version, the percentages of T1 – T4 values was somewhere between what had been predicted as a Day 2 or Day 3 mix as shown in Table 8:

T2(%)	T3(%)	T1(%)	T4(%)
68%	20.1%	6.4%	5.5%

Table 8: Initial Instantiation T-rating Distributions

The overall scores for the areas was shown in Table 8, which seemed to indicate that although the number of measures was probably excessive, the performance could be

assessed numerically, and the effects of high or low grades could be readily seen. Artificially lowering any high importance area (using the JMET driven view) did cause the overall score of any OE or APE area to drop significantly whereas artificially raising scores for very low importance areas didn't have a large impact.

The large number of measures in this model did not afford insights which would become evident as the performance assessment tool was refined.

4.4.2 Major Combat Operations (MCO) Model

4.4.2.1 Opportunity presented by Fleet Synthetic Training (FST) 07-2

The Fleet Synthetic Training (FST) series of exercises (formerly Battle Group In Port Exercise (BGIE))was instituted in March 2004 as a method to train U.S. Navy Strike Groups in a realistic operational environment, with the ship's crews manning watch stations as would be expected while underway, but with all ships in port. This concept has evolved to include Joint players from all the U.S. Armed Forces as well as Coalition participants from the United Kingdom, Germany, Canada, and Australia (13). The focus of the exercises, as noted earlier in this paper, is primarily at the tactical level, emphasizing unit and group training. There is no reason that the Fleet Synthetic Training (FST) series of exercises cannot be extended to an operational level staff, in a manner similar to that currently employed by the U.S. Air Force in the Virtual Flag series of Distributed Mission Operations (DMO).

For the FST exercise conducted in July 2007, an opportunity to construct a template that applied to a Combined Forces Maritime Component Commander (CFMCC) for evaluation for validity was available and leveraged as an element of this dissertation. Elements of the Second Fleet staff played in the role of the CFMCC for FST – J (FST-Joint) 07-2, which comprised three different U.S. and Coalition Strike Groups as well as U.S. Air Force units and U.S. Army Air Defense Artillery units. Though the inclusion of the CFMCC was very late in exercise development, precluding an opportunity to fully evaluate the Staff's performance, there was an opportunity to develop a set of measures which would apply to the Major Combat Operations scenario being used to evaluate the

U.S. Navy Strike Groups and would be used again in a follow-on live, virtual, and constructive (LVC) exercise.

4.4.3 Use of Existing JMETs

The most expeditious method of formulating an assessment tool was to utilize JMETs that had existing standards. The JMETs that had been identified by both Organizational Element (OE) and by the Assess, Plan, Execute (APE) function in conjunction with the effort to baseline the various numbered Fleet staff capabilities, and became the logical choice.

The key Operational Tasks that would be necessary for a CFMCC to carry out in order to manage a force of the size and potential geographic dispersion as was available, as well as those necessary to coordinate the actions of other Component Commanders These key Operational Tasks were determined based upon the scenario training objectives. The selection primarily involved the higher level tasks at the two-digit level (e.g. OP 1.2 Conduct Operational Maneuver and Force Positioning (1) but there was a need to identify some specific subordinate tasks for specialized missions; such as OP 6.2.9 Coordinate Personnel Recovery (1) as well.

A master spreadsheet model which included all measures with existing standards had been built as a result of the work conducted for C2F. The initial approach for constructing the CFMCC assessment tool was to delete the measures that did not apply to the training objectives and stipulated scenario. This process required about eight hours to complete using the various capabilities within EXCEL™. Once the applicable JMETS had been chosen, they were segregated according to either OE or APE category. This segregation was further refined by consolidating all measures with common skill assignments (e.g. all JMET measures relating to Adaptability).

The ability to arrive at what appeared to be a workable framework of measures in a relatively short time seemed to be not only useful, but appeared to validate the earlier work to identify OE, APE, and skill assignments. A total of 1225 individual measures were required in this formulation to assess the performance of a CFMCC. This number seemed a bit larger than the totals the author had encountered in previous strike group

experiences, but since the required JMETs appeared to be adequately covered, and there would not be an effort to assess performance, this did not seem to be a matter of concern. A total of 1609 measures was available to assess all the JMETs; therefore using 1225 measures seemed appropriate since Major Combat Operations necessarily includes a very large number of those measures.

4.4.4 Issues with Multiple Assignments

A review of this initial MCO formulation was performed by the staff of Tactical Training Group Atlantic (TTGL) who is the Executive Agent for conducting the Atlantic Fleet FST exercise series. Two of the senior personnel on that staff who were very familiar with the scenario, the assessment criteria expected for the various strike groups, and who had routinely participated in assessing strike groups for all previous FST events, agreed to review the formulation developed for the MCO template.

The initial reaction of the two reviewers was that there seemed to be too many measures for a group of 20 to 25 potential exercise observers to either effectively or accurately evaluate. They noted that a typical observer in FST exercises had a total of approximately 20 measures over the course of a 4 day exercise, and that many of these were repetitive which permitted properly assessing trends in critical areas such as planning or current operations. Using that possible construct, the number should have been nearer to 400 – 500 measures.

Both reviewers also noted that upon inspection, there were many repetitive measures and that this approach could cause unnecessary conflict. The conflicts would arise when similar measures were evaluated by different OE or APE areas and these differences would require adjudication. As a matter of practicality, it was also both time consuming and inefficient to have similar quantities entered in multiple areas where a single entry would permit feedback to the training audience on a daily basis, which makes the feedback more effective as noted by McGinnis and Stone (8) as well as Sloman (10).

A review of the initial model was conducted on the basis of these comments. Using the Auto Filter capability resident in the EXCEL™ spreadsheet toolbar, similar

measures were examined in multiple spreadsheets. At first it was unclear why so many duplicates existed, though it was immediately evident that there were repetitive measures in both approaches.

Two reasons eventually became evident for the duplication. The first was due to the numbering system for all the measures required for entry into the RATIONAL REQUISITE PRO™ database. Each measure requires a unique identifier, which would be desirable for any database application, but had led to duplication because of the initial formulation discussed in section 4.1.3.2. The choice to have multiple assignment possibilities for either the OE or APE had led the person assisting in setting up the RATIONAL REQUISITE PRO™ data base to assign numbers in series based upon OE assignment. Thus it was common for a given JMET measure to have numbers in multiple (sometimes as many as three) OE categories. The reason for multiple assignments remained valid, but the need to limit the possibilities prior to formulating an assessment template had not been anticipated.

The overall approach to formulating the assessment tool was judged to be appropriate, and if the duplicative measures could be eliminated, the reviewers were satisfied that the final version would be suitable for assessing a CFMCC and could be used during the investigation of costs and ROI. Both reviewers felt that there were JMETs that existed in the complete UJTL that should be added in the assessment tool, even if standards would need to be devised to more accurately reflect the required capabilities of a CFMCC. They also stated that if the assessment tool (and more correctly) the JMETs identified were considered at the outset of any specific training exercise, then it would not be difficult to construct a more appropriate and useable tool from the initial attempt.

The issue of multiple assignments was mitigated by selecting both the most appropriate OE and APE for each measure. Once that was done, the measures were segregated on this basis and a second check to reduce any remaining duplicate measures was conducted using the Auto Filter capability resident in EXCEL. This approach resulted in a significantly reduced number of measures, 692, which is still probably somewhat extensive for any assessment team, but which could easily be reduced if this

tool were reviewed in the manner discussed in the previous paragraph. Ultimately, the issue of multiple assignments was completely avoided by the Operational Level staff using unique Training Objective numbers for each measure.

4.4.5 Prediction Model.

An issue that consistently caused concern and elicited comment is, the question of "how this compares to any previous assessment process." The simple answer is there are no existing numerical assessments of operational level performance. The assessments conducted by the Observer Trainer teams from Joint Forces Command heretofore (and will probably remain for some time to come) support the preparations of the CJTF and his staff for deployment. These assessments do highlight best practices as well as advising the Commander and his staff of areas where additional emphasis or training could be helpful. However, the ultimate assessment and responsibility for the reporting of readiness is left (and even with the methodology being developed herein will remain) with the Commander. His judgment of his and his staff's performance is influenced by the observations of the JFCOM team, but these observations are not directive in nature nor are these observations quantified in terms of measures of performance, etc.

So, with the ever increasing requirements of DRRS, how does one translate readiness reporting beyond a simple color scale or a Pass/Fail system? One compelling answer is by utilizing scales which more accurately reflect readiness or training. These can be readily adapted to describe "best practices" which would correspond to T-1 (better than average), T-2 (operationally ready), T-3 (requiring additional training or in some usages "Partially Trained"), or T-4 (unsatisfactory). Those types of scales exist for the Tactical level of training for U.S. Navy Strike Groups, but are not currently applied to operational level Commands.

Adopting a T-1 to T-4 scale will not inherently provide a historical trend until this methodology has been employed for several CJTF operational level events, and probably not until there have been several years of data collected. There are approximately five to ten events per year which could be candidates for such as system. A year's worth of data would be a start, but two or three years would probably be more useful to analyze trends,

especially since some of the CJTF staffs would be repeating the initial certification and sustainment cycles.

The preceding discussion establishes what is needed and how it could be used, but does not answer the initial question, how does the performance portrayed compare to any previous assessments? To expand on this discussion, a reasonable way to answer that question is the use of data that has already been collected from similar Mission Essential Tasks from staffs that had to carry out similar functions. Although the Commanders for Carrier Strike Groups are not of the same seniority as a CJTF, they and their staffs must carry out many of the planning, communications, assessment, and information operations functions required of a CJTF staff. If such data could be obtained, and there were a sufficient number of instances where similar METs had been evaluated in the three years that Strike Groups had been undergoing certification, this could provide at least one basis for performance comparison.

Unfortunately, the data which relates to the performance of Strike Group staffs, especially that data which relates directly to the JMETs of interest in this research is not easily accessed. The issue concerning examination is not classification, although it is classified as CONFIDENTIAL, but rather the method of organization of the data. The JMETS of concern must be related to similar Navy Mission Essential Tasks, which is not difficult. The difficulty is the very cumbersome method of storage and the inability to relate the observed values to standards and to specific days within an exercise. All of this could be overcome given sufficient time and software products, but the effort was not considered practical for this research.

The next best method to represent training achievement, especially if daily trends were either desired or actually observed, is to build an initial formulation based upon experience. The exercises conducted by TTGL typically cover a four day period, which is the same period as designated for the CJTF exercise being discussed within this paper, so it appears that perhaps the trends that had been noted would at least provide a compelling baseline for comparison.

The staffs for all Atlantic Fleet Strike Groups since 2003 have been certified for deployment, which means they have achieved at least a T-2, or operationally ready rating. It is common for the groups to start training with a significant number of measures at some value other than T-2. There are a few that are T-1, but the majority of the non-T-2 ratings are T-3, and some are even T-4. A construct of 60% T-2, with 20% T-3, and then a simple split of 10% each for T-1 and T-4 was logical. The progression was then set to permit an increasing capability by increasing the T-2 rating by 10% each day and proportionally decreasing the T-3, and T-4 ratings. If such a progression was followed for a four day period the results would seem to follow the values shown in Table 9 below:

	T2(%)	T3(%)	T1(%)	T4(%)	Overall
Day 1	60	20	10	10	.7550
Day 2	70	15	7.5	7.5	.7788
Day 3	80	10	5	5	.8025
Day 4	90	5	2.5	2.5	.8625

Table 9: Projected Performance Trends

This approach also seemed to be fairly simple to implement on the initial MCO models, but it did not allow for the fact that staffs had a "learning curve" that did not follow the normal distribution as discussed above. In almost all cases that could be remembered by the researcher and those who have conducted Strike Group training, it was unusual for any group that could achieve a T-1 rating in any measure to have that measure degrade to a lower value as training progressed. There was a steady improvement noted in all Strike Groups especially in the initial T-4 and T-3 ratings since these were more heavily emphasized by the assessment staff and the staff being trained. There was also a normal increase in T-1 ratings throughout the training period, so the final distribution was not normal, and this was much more difficult to model.

Two other possible final values for T-rating distributions were attempted in the initial MCO and CJTF certification models. These involved final values (starting as projected above and then concluding with the percentages shown below) for T2 at either 70 or 75% with the corresponding percentages for T-4 at only 5% and then either a 15% or 10% achievement of T-1 and T-3 ratings accordingly. (70(T2)-15(T3)-10(T1)-5(T4) or 75-10-10-5) The expected linear predictions are shown in Tables 10 and 11. The values which the methodology achieved are also depicted. The actual results noted for the various applications other than the MCO test model are discussed in the appropriate sections that follow.

	T2(%)	T3(%)	T1(%)	T4(%)	Overall
Predicted	70	15	10	5	.8025
Observed					.7938

Table 10: First Change to Performance Prediction[24]

[24] Observed refers to the overall performance value for the Assess category with the same percentabe of T-rating assignments as predicted with the results from the calculations in the performance assessment tool.

	T2(%)	T3(%)	T1(%)	T4(%)	Overall
Predicted	75	10	10	5	.8075
Observed					.7725

Table 11: Second Change to Performance Prediction

4.4.6 Noble Resolve Experiment Series Model

4.4.6.1 Opportunity presented by Noble Resolve 07-2

The Noble Resolve Series of experiments is carried out by the J-9 organization at Joint Forces Command, which is responsible for Joint Concept Development and Experimentation. One particular experimentation area of interest is the development of architectures that can help support Emergency Preparedness for Federal organizations such as the Department of Homeland Security (DHS) and various state level Emergency Operations Centers and training events. During the month of August, the states of Oregon and Hawaii, as well as the territory of Guam conducted an exercise to test the abilities of each organization's Emergency Response teams. J-9 provided scenario background as well as various architecture and network tools to assist in improving connectivity as well as to provide a means to increase the situational awareness of each organization's Emergency Operations Center team as well as those government and first response agencies whose capabilities and personnel would be called upon to assist during an emergency response scenario whether caused by a natural occurrence or as the result of a terrorist action.

The ability to assess the performance of the exercise participants, the technologies being tested, or the utility of the scenario and exercise control team is not carried out in an objective manner with published standards; rather, it is based on daily after action briefs and a final After Action Review which includes the leadership from each participant. The items identified as those that should be studied for improvement or to be categorized as lessons learned are discussed and reported for action by the appropriate agency. There is no formal method of capturing the end state proficiency of any

organization or identification of operational shortcomings which can be addressed through further training and emphasized in later exercises.

The possibility of building a framework which might be used to provide insight to the issues of proficiency, exercise shortfalls, technology gaps, or scenario development was welcomed by the Noble Resolve Series Coordinator. No formal assessment of any staff involved would be conducted, but the ability to review the applicability of a performance assessment tool designed specifically for this sort of exercise was desired. The experiment also covered a sufficient period of time (four days) and could be monitored remotely (from within the J-9 Joint Operations Center) that at least a rudimentary assessment for two successive days of scenario play could be accomplished to provide a sense of the effectiveness of the exercise. The costing aspects that are associated with this effort will be discussed in a later section, but those were also a desired facet of the experiment.

4.4.7 Use of Existing JMETS

As was the case with the Major Combat Operations performance assessment tool there is no current formalized assessment methodology either available or being used to support the Noble Resolve Experiment series. The Federal Government has very recently issued a Target Capabilities List subtitled, "A companion to the National Preparedness Guidelines" to, "serve as a framework to guide operational readiness planning, priority-setting, and program implementation at all levels of government (64)." The document was unavailable when the request to support Noble Resolve was made. Further, there was not time, even if the document had been available, to tailor the guidelines for each state or territory involved.

The basic performance assessment tool for an operational level CJTF certification exercise had been formulated to support an exercise which would follow this experiment by about two weeks. Since the Foreign Humanitarian Assistance and Disaster Relief requirements identified as necessary components of the certification process were similar to the requirements that a state Emergency Operations staff would require, and the purely military aspects of the tool did not appear to be excessive (e.g. measures designed to

assess the ability to rapidly strike highly mobile targets or utilize large amounts of imagery for decision making) the "as formulated" performance tool was offered and accepted as the baseline performance assessment tool for the experiment.

4.4.8 Changes due to Homeland Security Scenario

The performance assessment tool originally accepted as the basis for developing a performance assessment tool for experimentation in connection with the Noble Resolve series was not as extensive as the first Major Combat Operations model but still contained over 700 distinct measures. As the scenario progressed, it became obvious that at several differences between military applications and an application that would more appropriately apply to either a Defense in Support of Civil Authority (DSCA) or an application tailored to an entity that was relying on some form of military assistance must be considered.

First, even in operations such as Small Scale Contingency Operations or Stability Operations, there is considerable use of Intelligence data to assist in various types of targeting or specialized mission preparation. This type of intelligence is not relevant for the vast majority of operations carried out during a disaster scenario. One instance where this type of information could be utilized effectively would be in an Anti-Terrorist action, but then very specialized military and civil law enforcement units would be involved and these generally have the types of data and support required. Any additional information required can be arranged outside of the channels being used by state or other agency Operations Centers. Observing the scenario unfold over a period of three days enabled the JMETS relating to this type of intelligence data to be deleted.

Significant numbers of JMETs involving the execution or monitoring of specific military operations, the integration of plans and operations that related to specific military operations, and the control of specific military operations were included in the originally offered assessment tool. As discussed previously, the reason for these inclusions was because although the mission requirements that were deemed essential by the CJTF to attain certification included some very specific military operations such as surface ship board and search operations, or the ability to develop branch and sequel plans in response

to strike operations, none of these types of operations apply in a Disaster Recovery scenario or even a joint military and civil authority action against terrorists.[25]

One interesting aspect of Noble Resolve 07-2 was that the JMETs that had been forecast to apply to the CJTF scenario were also applicable (with only one exception) to the Disaster Recovery scenario used for Noble Resolve. The necessity of Logistics support to successfully carry out any large scale operation is critical, the primary difference being what is being transported or stockpiled for mission accomplishment.

As a basis for developing a cost model, it was necessary to establish the core set of JMETs that would apply to the Noble Resolve scenario. An additional consideration for the development of the core set was to ensure that these could be used for later Department of Homeland Security (DHS) scenarios with the eventual goal of linking the JMETs identified to the measures that are similar in the Homeland Security Universal Task List (64).

The core set of JMETs agreed to is listed below:

OP 1.1 Conduct Operational Movement

OP 4 Provide Operational Logistics and Personnel Support

OP 4.7 Provide Politico-Military Support to Other Nations, Groups, and Government Agencies

OP 5.1 Acquire and Communicate Operational Level Information and Maintain Status

OP 5.2 Assess Operational Situation

OP 5.5 Establish, Organize, and Operate a Joint Force Headquarters

---

[25] Observed refers to the overall performance value for the Assess category with the same percentage of T-rating assignments as predicted with the results from the calculations in the performance assessment tool.

OP 5.7 Coordinate and Integrate Joint/Multinational and Interagency Support (1)

The listings above are the JMETs that serve as the basis for all other supporting tasks. This approach is similar to the approach which was used by the Operational Level staff to develop the performance assessment tool for FA PANAMAX 07. The eventual number of JMETs that did not apply to the scenario was 166. This made the overall total of measures requiring observation for assessment of the experiment a much more reasonable 544.[26]

There were a number of measures worded in a military context that seemed appropriate to retain. An example of such a measure would be the necessity to identify critical C4I architecture nodes identified in OPLAN (OP 5.3.8 M5) (1) and the requirement to establish a secure and redundant communications system that ensures connectivity between critical agencies (ComC 1.4.2) (64) Although the HLS UTL was not published at the time of Noble Resolve, there were 41 measures from the UJTL that seemed as though they would apply in the Noble Resolve scenario. These measures were retained in the performance assessment tool.

4.4.9 Development of Radioactive Dispersal Device JMETS

During the course of Noble Resolve 07-2 the fact that the Territory of Guam wanted to experiment with assessing the territory's Emergency Response capability when faced with the detonation of a Radioactive Dispersal Device (RDD). This was important because the Territory would be conducting the same scenario, but in a LVC construct in October during Exercise TOPOFF 04, and needed to have some indication of how existing plans and Emergency Response and communication capabilities might perform.

---

25 Guideline that is currently followed by most U.S. Navy assessment teams is that any single assessor can be assigned between 20-25 measures for evaluation. This total also assumes that over an eight to twelve hour shift the total can be comfortably assessed and recorded. Thus, the number cited would require approximately 25 assessors to effectively evaluate the experiment.

The personnel within JFCOM J-9 were aware of the plan to test the Territory's response to a RDD, but had not requested specific measures that might be applied for assessment before the experiment began. A request was made to develop an initial set of measures for review by both J-9 personnel and Emergency Response Coordinators in the Territory of Guam. Since no specific details of the scenario were known beforehand, a generic set of measures would at least permit an assessment of the typical types of actions that would be expected from an Emergency Operations Center, the leadership team, medical teams, and outside agencies.

The portion of the UJTL that applies to Chemical, Biological, Radiological Nuclear or high-yield Explosive (CBRNE) (1) is OP 7 and the subchapters within. A review of the existing measures resulted in the selection of 49 measures that could apply to a RDD event.

An important distinction from all previous model constructs was that very few of the measures within OP 7 had been addressed by any of the three organizations which had developed standards for the UJTL. The need for measures necessitated several actions be undertaken for those not previously addressed to permit incorporation into the assessment tool. First, each measure needed to be weighted as had been done with all previous measures. Each also required a skill assignment and assignment to the appropriate OE and APE category. The final step was to determine a range of standards for assessment. These would also necessarily require the same types of attributes, time, percentage, number of instances, or perhaps a binary Yes/No attribute in some cases.

A total of 47 JMET standards were developed using the descriptions from the measures as the guidelines for standard development.[27] Even with the descriptions for

---

[27] The descriptions do lead those devising standards some idea of scale such as OP7.2 Measure 15, which discusses the instances when medical facilities are not able to treat contaminated casualties. This should be competency of most medical facilities, so the number of instances for which care cannot be provided should be low, but not zero owing to potential damage to some facilities or being otherwise inaccessible.

the necessary criteria for successful achievement being available, substantial reliance on previous experience with nuclear power applications enabled the researcher to more easily make the necessary assignments. The standards proposed were reviewed and concurred with by J-9 (personnel and then reviewed again for applicability by personnel from the Territory of Guam during Top Official 04, where the final number of measures, 47 was accepted.

The developed measures were incorporated into the appropriate sections of the performance assessment tool and were available for data input during the daily After Action Review (AAR) sessions as well as during the scenario play. The entire process required a bit more than two hours from initiation to completion, but that was aided immeasurably by the researcher's previous experience and familiarity with the methodology and performance assessment tool.

As a result of the work to include the RDD event, three additional "core" JMETs were added to the listing shown above. These would also be utilized for the cost model. The three additional JMETs are:

OP 7 Counter CBRNE Weapons in the Joint Operations Area

OP 7.3 Coordinate Passive CBRNE Defense in the Joint Operations Area

OP 7.4 Coordinate Consequence Management (CM) in Joint Operations Areas (JOAs) (1)

4.4.10 Assessing NR 07 From Daily After Action Reviews

As noted from the beginning of the description of the construction of the performance assessment tool for Noble Resolve, there had been no pre-existing agreement with any of the states or territories or with the J-9 to establish assessment teams or to evaluate the experiment in terms of performance for any of the states, agencies, or organizations involved. Yet, there was a desire to at least gain an understanding of the potential to carry out such an evaluation, even if that was done remotely for all but the J-9 personnel to gain an understanding of the capability of the architecture to provide the proper connectivity for decision making as well as to assess

the effectiveness of some of the situational awareness tools such as Event News Network (ENN) that were being employed to increase the realism of the disaster scenarios.[28]

The experiment was debriefed daily in a Video Teleconference Format as well as in a VTC format with the leadership teams from each state, agency, or organization participating on the last day of the experiment. The major issues concerning each participant were discussed, and those that were causing either significant situational awareness issues or an inability to maintain a clear operational picture were forwarded to the appropriate support team for resolution.

The majority of issues with the experiment dealt with an inability to access information uniformly (some locations could access the network being used to carry the voice, data, and imagery while others could not) or the lack of access to specific classified data that was relevant to the scenario. These comments reflected delays in decisions or some repetitive requests for information during the day's events, which would have caused lower evaluations in some measures. An example of both a daily briefing slide and notes capturing some of the verbal comments are shown in Appendices D and E. The daily AAR generally reinforced the observations made by the researcher with the various communications systems available in the J-9 operations spaces, and was also generally consistent with the thoughts of the J-9 personnel conducting the experiment.

---

[28] A major earthquake with resultant damage and the subsequent failure of a major dam which would also impact the city of Portland were the subject of "footage" and "live" reports to aid in scenario play.

4.5 **Cost Modeling for the CJTF Certification Event**

4.5.1 Ability to Capture Costs for CJTF Certification Event

The work done on the two preliminary instances for costs mirrored the efforts required to formulate the performance assessment tool. The Operational Level Certification presented a much greater potential to develop the cost model and use it for purposes other than merely comparison between events.

The CJTF had expressed a desire to have a numerical method of assessing the performance of the staff during the certification exercise. The fact that the performance was tied to doctrine, in this case the UJTL, as well as having specific numerical results that could be used for follow-up training and trend analysis was appealing to the CJTF. The Commander also stipulated that the staff use well defined processes, which could be easily written on single sheets of paper, similar to the "knee boards" used in aviation, whenever possible. These objectives were met, but a greater concern of the Commander was the ability to view how much the event cost.

Since this was the very first Operational Level certification, the Commander felt that there would be many costs that would fall into the One-Time/Event Specific category. There was also confidence that there would be some Indirect costs that would be incurred due to the staff configuration[29] chosen to execute the certification. It was the Commander's belief that unless ALL costs were captured and categorized, that the lessons learned from this event, and any possible advantages that might be gained by other similar commands would be either lost or repeated, and either course was unacceptable. The Commander also believed that a close examination of the costs would provide the basis to defend the costs to establish the CJTF capability, but it would also

---

[29] The staff was split between two geographically distinct operating points, as could very well be expected in the missions specified by the COCOM. Past experience from Hurricane Katrina had shown this model to be viable, but the actual cost impacts and operational limitations had not been fully explored.

permit those costs to be avoided, either in totality or at least substantially for that staff and any others in a similar situation.

The very great concern over costs shown by the Commander made working with the Staff to capture the costs quite easy. The staff was very open to gathering all data and was a very willing partner in assigning the costs to the appropriate cost categories as well as to the most appropriate JMET.

4.5.2 Cost Assignment to JMET Process

In contrast to the previous use cases for costs, there was no question as to which JMETs would be chosen for cost assignment. These were already delineated in section 4.4.2 and will not be repeated here. One aspect of the JMETs briefly mentioned in the earlier section was that the Training Objectives chosen to represent both OP 1.5 and OP 6.5 were identical to those chosen for OP 1.2 and OP 6.2 respectively. There was no need to assign costs to all four JMETs; only the primary OP 1 or OP 6 category would be used.

As was found in the JMET assignments and in concert with the attitude of the staff to assign costs to the major cost categories, the staff readily assisted in the assignment of costs to specific JMETs. This process took less than two hours to complete, and was agreed to in each assignment, so no review or re-submission for validation was required. The cost breakouts by JMET are depicted in Appendix I and follow the same methodology as was applied to both the MCO and Noble Resolve 07-2 cost models. The effort put into a more specific parsing for personnel done in the Noble Resolve 07-2 model was very helpful in rapidly assigning costs.

4.5.3 Utility of Model for Future Events and Other Commands

The costing effort executed by the Operational Level Commander has been projected to serve various functions. The first use is to trend costs for certification events so that budget forecasts for certification events and training required to support the certification process can be accurate. The Command must submit cost estimates to the

COCOM, to JFCOM, and to FFC for each event, although the use of these costs differs for each of those commands. The COCOM requires the costs to project the costs for each of the subordinate commands to maintain readiness against specified missions. The mission sets could also evolve and the ability to add JMETs with associated costs would be very useful.

JFCOM uses the costs to project exercise support in terms of scenario generation, facilities, and additional personnel both military and civilian to either execute the scenario or to assist in the exercise. The potential for personnel reductions, reuse of equipment or scenarios, or reduction in footprint are all significant cost factors in exercise planning for both JFCOM and any Operational Commands.

FFC uses the costs to estimate the additional systems, networks, hardware costs, or additional facility costs required for exercise (and in this case certification) execution. The ability to very specifically categorize the costs as well as the ability to identify the JMETs which the costs are attributed to is significant since FFC also funds similar operational level commands and decisions regarding funding for similar mission requirements can be made where appropriate.

One of the areas discussed above, expanding mission requirements is something that also intrigued the Operational Level Commander. There are some scenarios within the COCOM area of responsibility which might require additional JMETs, and having a methodology to assign costs to specific mission sets could provide insight and accuracy to budget forecasting.

The very large One-Time/Single Event and Indirect costs associated with this initial certification event were a concern of the Commander. These costs were deemed as necessary, but there was always the possibility that they could be construed as extravagant. The ability to assign the costs specifically to categories and to training that was accomplished at T2 provides the means to refute the assertion that money was ill spent. The knowledge of these costs also permits avoiding some or all of them for future events, so it is possible to demonstrate potential cost avoidance and then actually show

the degree to which that cost avoidance was achieved. This type of demonstration is only capable if the costs can be related directly to the training accomplished at a given level in both instances which can now be effectively shown.

The final utility of this methodology for any operational level Commander concerns the level of readiness of the personnel who are being certified. In this particular instance, the staff was well trained, knew when the event would occur, and had time to ensure that all preparations necessary in terms of facilities, assistance, and systems were in place. It is not hard to envision very different circumstances where there are far fewer qualified personnel, the mission requirements are markedly different, or preparations are very time compressed. The costs projected for the "normal" case will not be useful, and the overall costs required to prepare and deploy the staff could be significantly greater. As long as the basis for the costs for categories and JMETs are captured in the same formal methodology, the differences in costs will be quantifiable and justification of the resource expense should be more easily accepted.

4.6 Lessons Learned from Previous Cost Modeling Efforts

4.6.1.1 Breakdown Approach for Comparing Performance to Costs for the Major Combat Operations (MCO) case

The development of the cost models was accelerated by the offer from Fleet Forces Command (FFC) of the cost data from two different FST events as well as the cost data from Joint Force Exercise 06-2. The FST data permitted viewing two very useful cases, comparison of a LVC event to a completely synthetic event and the comparison of two completely synthetic events in two successive years.

The concept of using four major categories of costs, Direct, Personnel, Indirect, and One-Time/Event Specific was validated in developing the MCO case cost model. The data available was categorized by either a Live or Synthetic category and detailed all costs reported to Fleet Forces Command for services and personnel. Two categories that were not captured, and that were not supported by the reported data were either Indirect

costs or One-Time/Event Specific Costs. One of the problems with any costing strategy is the ability and willingness to capture all relevant costs. Though there may have been costs that would have been applicable to either of the missing categories, these were not pursued and could not be considered in any development work.

4.6.2 Cost Assignment to Major Cost Areas

The information supplied by FFC was assigned into the two categories for which data had been collected. The ability to capture specific costs regarding the fuel costs, fuel burn rates, pier services (required for docking and undocking), range services (for targets used during the live exercise) and even environmental costs is formalized at FFC. Additionally, the ability to capture synthetic costs such as networking costs is also formalized. The various types of personnel costs, including travel, surge labor support for exercise execution, and the costs incurred to ensure connectivity are also formalized. Once the costs had been categorized for the FY 06 events, the data was returned to FFC for review to validate the appropriateness of the assignments to the major cost areas.

4.6.3 Data Available for Three Events

The most recent FST event concluded in mid-July 2007. Collecting this data was accomplished by early August. The data was easily categorized as the data for the FY 06 event had been in only an hour or so. The addition of this data to that already on hand provided the basis for the first JMET comparisons.

One issue that was discussed with FFC was the need to include port services costs to all ships while in port. While it is true that these costs must be budgeted for by FFC, it seemed inaccurate to include them in the synthetic training model because the costs, which include sewage, fresh water, steam, and power are charged whether an exercise was being conducted aboard the ships or not. All costs relating to connectivity, personnel accommodation, and scenario specific devices are already being accounted for, so it was agreed that these costs would be deleted from consideration.

4.6.4 JMET Cost Assignment

The performance assessment model was amended to include four additional worksheets. These sheets were named for each of the cost categories. The issue was how to assign costs to specific training events as discussed in section 3.3.1 The costs didn't need to be applied to a specific event since there were multiple synthetic events and the method used to capture the costs was repeated from the FY 06 event to the FY 07 event. Further investigation of the costs made it seem possible to assign costs down to a three digit JMET, as would be applicable for Theater Missile Defense covered in OP 6.1.5 (Conduct Joint Operations Area Missile Defense) (1). The use of the U.S. Army Patriot Battery simulators, the Airborne Warning and Control System (AWACS) (67) simulators and personnel, and the Combined Air Operations Center (CAOC) personnel were all either specifically or at least partially used to achieve the measures within OP 6.1.5. The issue with such a limiting assignment would be that either another assignment for the AWACS personnel and CAOC personnel would need to be made because these personnel also have responsibilities within the Strike and Air Defense areas, or that there would be redundancy in the costing which would not reflect costs accurately, much as redundancy in the JMETs accomplished was inaccurate, and confusing.

There was still the possibility that costs might need to be assigned to either specific skills or functions. Again, since the costs were so well categorized, the use of either of these categories seemed impractical. Other costs were capable of being assigned to JMETs such as OP 6.2 (Provide Operational Force Protection), OP 1.2 (Conduct Operational Maneuver and Force Positioning), or OP 3.2 (Attack Operational Targets) so the most logical and consistent method of assigning costs was to assign them to JMETs as originally intended.

The assignments did require some decisions as indicated earlier. The use of aircraft and the Command and Control (C2) (67)capabilities required for employment needed to also be factored into the costs for these assets in addition to the specific OP 6.1 or other JMETs that were executed during the scenario. In order to do this in a consistent manner, the amount of scenario time devoted to activities such as strike, theater air

defense (for AWACS and other multi-use air assets) or C2 was used as the method to divide the costs and assign them. This could be done in either a LVC or synthetic environment, and would permit comparison on an equitable basis.

The only mismatches in costs, which could be used in a macroscopic view, but would not represent a one-to-one comparison were the costs incurred in getting the ships underway and to the operational area. The original cost assignments had been made such that the costs for pier services were assigned to OP 1.2. Since ships were in the theater concerned during the synthetic event and the costs for normal in port services had been judged as inapplicable, these costs were considered in two separate cases which will be further discussed in the results. The costing categorization and JMET assignments are shown in their entirety in Appendix G.

4.6.5 Cost Modeling for Noble Resolve (NR) Model

4.6.5.1 Ability to Capture Costs for NR Series

The Noble Resolve series presented another opportunity to capture costs for comparison between either successive events or to view some portions of the costs in a LVC vs. synthetic manner. Additionally, since the J-9 directorate of JFCOM uses the Noble Resolve series for experimentation, the establishment of a "core" set of JMETs, even if these are only accomplished versus evaluated for specific performance affords the opportunity to view costs against similar JMET assignments. This is a similar view to that proposed for the J-7 directorate of JFCOM.

The costs for the experimentation are shared with other agencies to a much greater degree than in the case of either USFFC or an Operational Level Commander and the supporting Combatant Commander (COCOM). An excellent example of costs that were not required to be borne by J-9 to support Noble Resolve was the use of the Integrated Common Analytical Viewer (ICAV) which is a tool being developed by the Department of Homeland Security (DHS) as the primary Common Operational Picture (COP) situational awareness tool. There was no rental, installation, or networking costs incurred by either the J-9 Directorate or JFCOM to use this capability. The functionality

that ICAV supported would normally be carried over the Joint Training and Experimentation Network (JTEN) which does incur some networking costs.

No external Modeling and Simulation support was required because both DHS and NORTHERN COMMAND (NORTHCOM) used the Joint Semi-Automatic Forces (JSAF) capability to generate event simulations and stimulations as well as other modeling and simulation capabilities already resident within JFCOM.

There were no travel costs associated with J-9 since there were no experiment observers on station and any observers or support personnel from the participating agencies were funded by the parent organization.

The methodology used to assign costs to JMETs for Noble Resolve was similar to that developed for the MCO model. As will be discussed for the Operational Level staff exercise and certification, the scenario for Noble Resolve lends itself to a core set of JMETs. The core set of JMETs agreed to for the Noble Resolve was listed in Section 4.4.3.

The costs were exclusively related to the Personnel Category and could be further specifically associated with either Exercise Labor or Surge Support. Exercise labor consisted of three sub-categories: 1) Scenario Development, 2) Scenario Planning, and 3) Game Control. Surge support consisted of Modeling and Simulation support.

The personnel from the J-9 assisted the researcher in the JMET assignment process by referring to both the Final After Action Review (70) and the NR 07-2 Final Report DRAFT (29 October) (71). These documents provided the basis for all the participating agency objectives, the types connectivity that were being used to maintain situational awareness and as the primary means of information exchange, and finally the overarching goals for the experiment.

The parsing of resources for each of the core JMETs was accomplished by considering the factors discussed in the preceding paragraph as well as by the effort that J-9 personnel estimated was required for each JMET in terms of scenario generation and

planning or scenario play (game control). The parsing and the relative breakouts for the two digit JMETs is shown in Table 12 below as well as in the cost sheet in Appendix I. The first column indicates the percentage of the overall cost (in either support category) that each OP was assigned. The second column indicates the percentage of the OP assignment that was assigned. There are two instances where a single JMET was used as well as a two digit JMET. Further, to reach the individual percentage that a specific JMET was allocated, the initial percentage column for the OP would be multiplied by the final column. So OP 5.1 would have a total resource allocation equal to 0.55 times 0.5 or 0.275 of either the Exercise Labor or Surge Labor support costs.

OP	%Assigned	Two Digit OP	%Assigned
1	5%	1.1	5%
4	10%	4	30%
		4.7	70%
5	55%	5.1	50%
		5.2	20%
		5.5	10%
		5.7	20%
7	30%	7	10.6%
		7.3	36.2%
		7.4	53.2%

Table 12: JMET Parsing Scheme for Noble Resolve 07-2

A review of the OP 7 (Coordinate CBRNE Weapons in the Joint Operations Area) (1) revealed that of the 47 JMETs used in both the NR 07-2 and TOPOFF 04 scenarios, five were not from either OP 7.3 or OP 7.4. The very low percentage contribution of these specific OPs (less than 5% each) caused the researcher to request that these be combined under OP 7 (which then comprised approximately 10% of the JMET measures used). The resources allocated to OP 7 activities were then divided by the number of JMETs observed in both scenarios. This allocation is clearly depicted in Table 12.

The parsing scheme devised was validated by JFCOM J-9 personnel, and is available as a baseline for future events.

4.6.5.2 Cooperative Cost Assignment Results

The personnel from J-9 were most enthusiastic about the prospects of viewing costs in terms of Noble Resolve JMET accomplishment. The ability for the staff to be an integral part of the cost assignment process to both the major cost categories and to specific JMETs was very appealing to J-9 since it provided an opportunity to use a developmental tool and retain personnel with the ability to use it in future events.

4.6.5.2.1 Basis for Future Event Cost Analysis

The nature of the experimentation events executed by the J-9 Directorate at JFCOM lends itself very readily to the formal costing methods and JMET alignments presented here. Though it may not be practical to assess organizational performance outside of Joint Forces Command, successful completion of experimentation events within the Noble Resolve series can still be framed in terms of JMETs. It is also possible that the Homeland Security Universal Task List (UTL) could be used as the basis for viewing the tasks completed successfully (as determined by the leadership of the participating organizations, similar to the manner that operational level Commanders use the services of the J-7 Directorate at JFCOM to determine event completion) for any specific event.

As with the baseline tasks now available to the Commander completing CJTF certification, a baseline set of tasks in either the UJTL or UTL can be established. Once established, assigning costs to these tasks will permit assessment of cost advantages for emerging technologies, scenario enhancements (such as the Exercise News Network, ENN), and networking services.

Another potential tool for the J-9 as well as any organization would be to view the JMET assigned costs as a percentage of the overall cost to gain perspective as to the costs expended for any single JMET for the gains or losses in readiness assessed.

4.7 Trend Model for Joint Forces Command (JFCOM)

4.7.1 Use of a Streamlined Approach to Identify JMETs Historically Observed

The ability to formulate a performance assessment tool and to identify the key JMETs necessary for exercise design, exercise/experiment scenario design, or performance assessment appeared to be translatable to efforts within the Joint Warfighting Center. There are several exercise programs overseen by the J-7 (Training and Exercise) directorate within JFCOM that are recurrent and could benefit from at least the ability to compare costs over time or by event to determine ROI.

A proposal was made to the C4I/Simulation director and the senior civilian deputy of J-7 that might provide at least an initial capability to assess the benefits of the Unified Endeavor (UE) series. This series supports the final training MRX (Mission Readiness Exercise) framework to various Service operational level staffs preparing for deployment to Operation Iraqi Freedom. While some form of assessment is included in these exercises, it is not based upon the same numerical system identified and used in section 4.4, but the ability of the deploying CJTF to report readiness for deployment is met within the construct of the UE series.

The proposal was that a historical view of the UE series be conducted looking at the most common set of JMETs reported as completed for UEs over an agreed to period of time. The report of completion would be taken as a T2 rating for any JMETs reported. No subordinate JMETs would be investigated unless those were specified by any of the reporting commands. There was also the distinct possibility that there might not be completely common sets of JMETs, in which case, constructs for specific events could be created, but these would each be in isolation.

The possibility that there could be little commonality between events caused some concern, but if the circumstances for deployment required such differences to exist, then it was more useful to recognize that these differences could be factors which tended to

have a major impact on costs than to disregard any effort to gain insight into the exercise program

4.7.2 Use of Streamlined Approach to Quantify JMETs for Specific Exercise Event

An exercise within the UE series was scheduled for the early October 2007 timeframe. As an adjunct to the above, the JMETs targeted for completion during this MRX were proposed to be compared with the historical models described in section 4.5.1 Several benefits of this approach appeared possible, but by reviewing the costs associated with the October event which had the same JMET assignments as previous events, cost trends that might provide insight into efficiencies gained over time by improved scenario development, reduction in travel, or implementation of better communications capabilities might be evident. The insights gained from this comparative review could benefit leadership but could also point out specific instances where costs could not be avoided due to very time compressed CJTF identification and assignment or because significant personnel replacement would require additional costs not encountered with a fully trained staff.

5 MODEL RESULTS

5.1 Model Results

The final model which was used for the Operational Level Commander Certification exercise will be discussed first, but there were two other use cases which provided significant insight and results. These will be discussed as well since these cases provided additional validity to the research.

5.2 Operational Level Commander Certification Exercise Model Results

5.2.1 Performance Assessment Results

All previous performance assessment models had been based upon assumptions of performance or a limited number of observer's inputs. This was also true of the measures utilized, the weighting factors, and the skill assignments. The discussion throughout section 4.4 clearly demonstrates the complete participation of the Operational Level staff in the process. Though some portions of the originally submitted performance assessment tool were ultimately included in the final version, this was because some aspects of the certification did not change from the originally stated mission requirements from the Combatant Commander.

5.2.1.1 Original FA PANAMAX Performance Assessment Tool

The version of the performance assessment tool submitted to the Operational Level staff for consideration was based on a set of mission requirements that did not include any significant combat operations, notably the absence of Strike Warfare requirements; thus, it had far fewer operational intelligence issues. The original Operational Missions included the following JMETs:

OP 1.1 Conduct Operational Movement

OP 4 Provide Operational Logistics and Personnel Support

OP 4.7 Provide Politico-Military Support to Other Nations, Groups and Government Agencies

OP 5.2 Assess Operational Situation

OP 5.5 Establish, Organize, and Operate a Joint Force Headquarters

OP 5.7 Coordinate and Integrate Joint/Multinational and Interagency Support

OP 6.2 Provide Protection for Operational Forces, Means, and Noncombatants (1)

Even though these were the primary JMETs, none could be accomplished in isolation, as demonstrated in all the other performance assessment tool constructs. A preliminary performance assessment tool addressing these mission areas was constructed with the goal to limit the number of measures requiring observation to the 20 – 25 per observer discussed in sections 4.2.2 and 4.4.4. The initial model consisted of 813 measures, which significantly exceeded the desired number of 500 measures. (There would be approximately 25 observers available with a wide variety of experience levels). The overall number was high, but there seemed to be an adequate coverage of the desired JMETs and it would simply be a manner of examining the proposal with the leader of each Organizational Element to determine what measures were sufficient to assess the performance of the staff. A byproduct of this examination would be the chance to validate the weighting assignments, skill assignments, and finally LDF factors.

While waiting on the opportunity to conduct each of the OE meetings, validation that the performance assessment tool did not contain any redundancies, that changes in observed values from day-to-day were consistent and were properly reflected, and that the AHP (JMET driven) values were consistent from day-to-day was undertaken. The results of the JMET T-rating distributions (noting that no T-1 values ever decreased, and that T-3 and T-4 ratings would diminish in a progression similar to the progression shown in Table 9) from day to day for the test FA PANAMAX performance assessment tool is depicted in Table 13.

	T1	T2	T3	T4
Day 1	0.0959	0.5904	0.2042	0.1095
Day 2	0.1439	0.8069	0.0455	0.0037
Day 3	0.1480	.0.7783	0.0703	0.0025
Day 4	0.1685	0.8007	0.0308	0.000

Table 13: FA PANAMAX T-rating progression

The results above appeared more realistic than those proposed in section 4.4.5, where the T-ratings followed a very predictable pattern, and since there was no event to compare this progression against, the results did reflect the ability to improve performance in a manner that had been observed by the researcher.

The formulation of four days of scenario "play" also allowed for trend data to be plotted against the initial performance prediction model in Table 10. The LDF factors were also calculated for each OE and APE approach. Figures 12 and 13 show the values for Assess, Plan, and Execute after applying the AHP factors (JMET driven) either using the Normalized averaging method or the Floating Weight averaging method. The numerical values for all the figures in the subsequent discussions are available in Appendix K. The agreement for the either APE approach is apparent in both magnitude and trend which was confirmation that for this large number of measures, even with a disparity between the total numbers of Assess measures (less than 100) and the Plan or Execute measures (significantly more than 200 each) the contributions of each were still influential enough to permit overall agreement between the aggregation of OE measures and the directly calculated APE values.

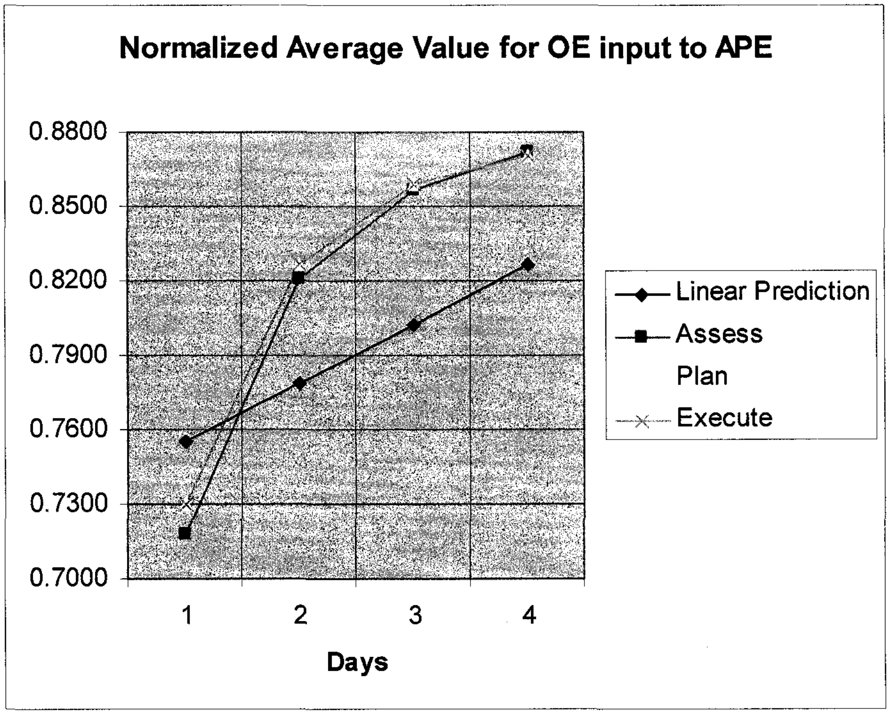

Figure 12: Normalized Average Values for OE Inputs to APE

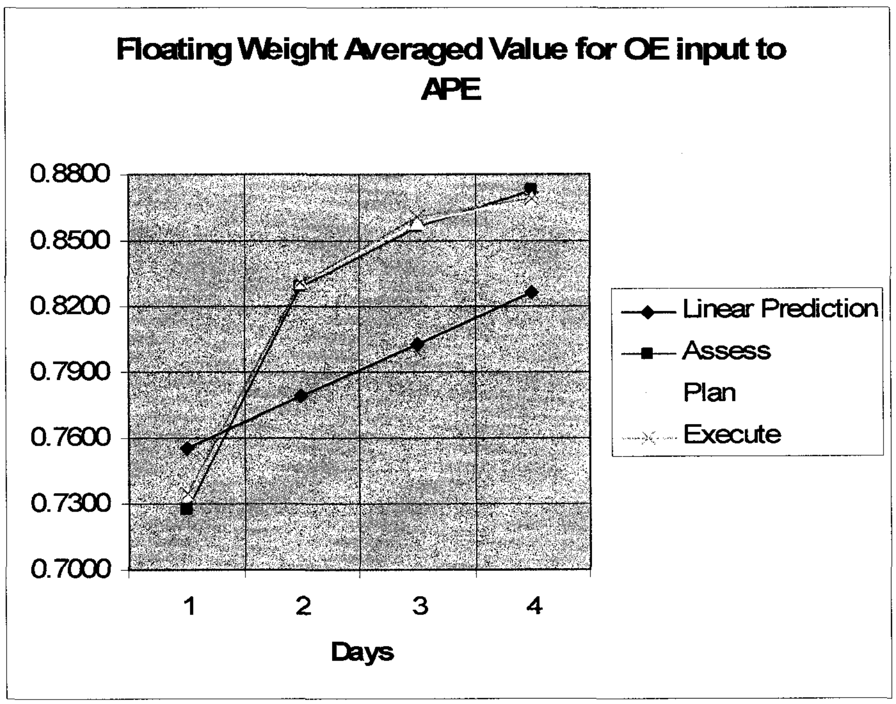

Figure 13: Floating Weight Averaged Values for OE Inputs to APE

Figures 14 and 15 depict the values (both Normalized and Floating Weight averaged) obtained when directly calculating APE. The trends are similar in both shape and value to the results obtained utilizing the AHP factors, which was encouraging since it appeared that either method could be used to obtain an APE value and that there was a definite linkage between the OE values and the APE values when the AHP factors were used.

The last set of results which were anticipated as a very definitive method of not only portraying APE performance, but as a very sensitive method for indicating substandard performance were the results for Logarithmic Driving Factor (LDF). The utility of the LDF approach when used to calculate performance values for OE applications had not been promising in earlier versions of the performance assessment model. The results for each OE area typically yielded very low results for LDF when all the subcomponents of each skill assignment area were multiplied together, but the possibility of aggregating the skill assignment areas to obtain a result for LDF similar to that which would be calculated from the APE subcomponent areas directly remained a potential method of relating OE to APE.. The results of both the aggregation method and the product of the OE areas within the APE context are shown in Figures 16 and 17:

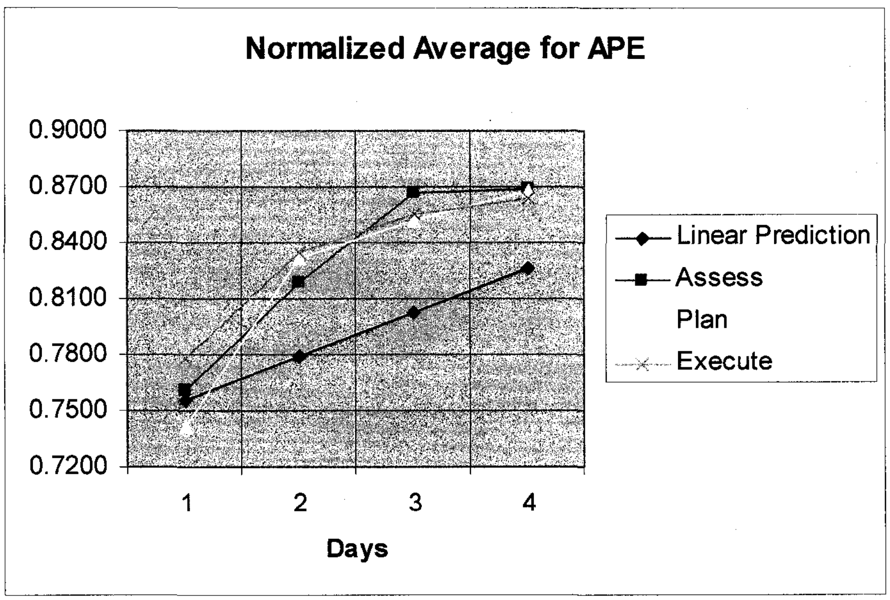

Figure 14: Normalized Average Values for APE

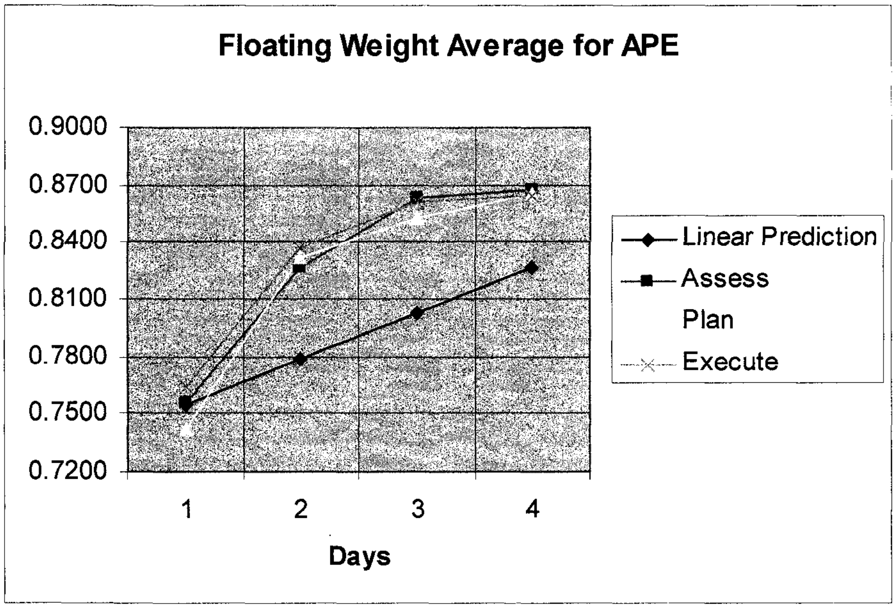

Figure 15: Floating Weight Average Values for APE

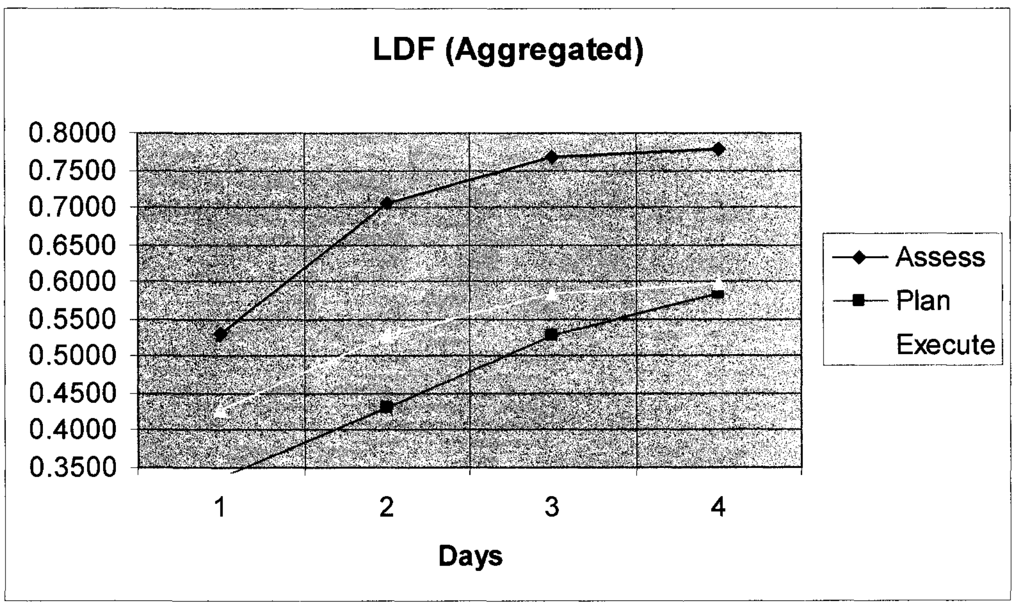

Figure 16: Aggregated LDF Values for APE

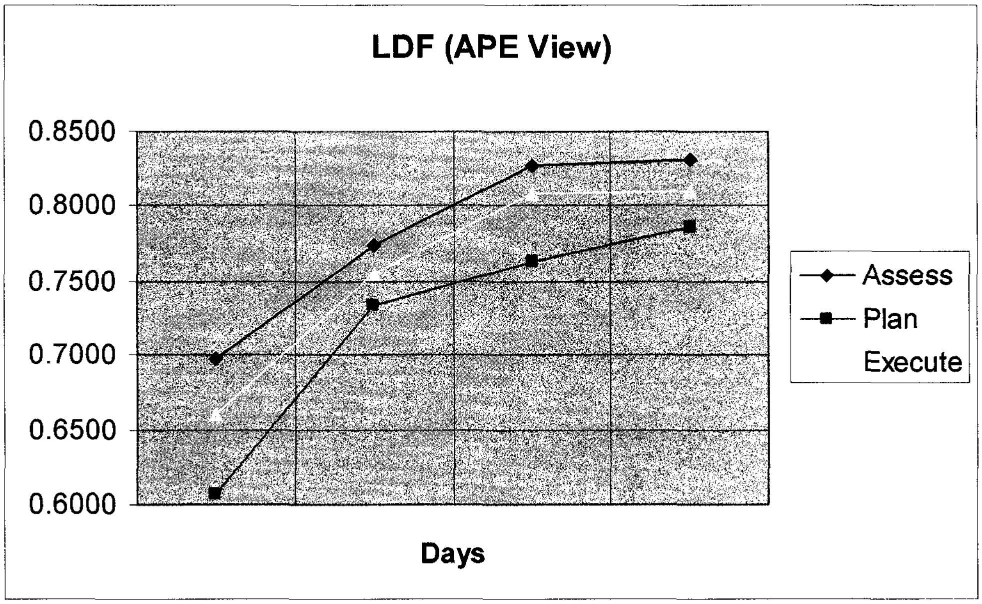

Figure 17: LDF Values for APE

The very noticeable difference between Figures 16 and 17 and the previous four figures is that though the shapes of the curves for the trends bear some resemblance, the numerical results, especially in the case of the aggregated LDF values, do not compare favorably with either averaging method. The directly calculated values for LDF, depicted in Figure 17, are better approximations for performance than the aggregated values, but these values still vary from the averaged values by at least 10%, which would not be useful to a commander. Only the values for Assess in the aggregated LDF approach approximate the values for Assess, Plan, or Execute in the other calculations. Although there were already two reliable methods for calculating performance values as well, what appeared to be a very reliable method for relating the OE values to APE, the potential benefits of the LDF method were the motivation behind continued development of this approach since a significant change(either improvement or decline) in performance within a key skill assignment area if utilizing the second level aggregation or within a given OE area in the APE calculations would readily be apparent. The ability to isolate the deficiency responsible for the degraded score would also be readily found which would permit a staff to decide on a remediation method or a process to improve

performance. The ability to isolate such deficiencies remains a tremendously powerful benefit of the APE cycle. The second level aggregation method had not been attempted in any earlier performance assessment model instantiations, but seemed to be a logical approach to more accurately depicting the APE performance in terms of OE performance than even a linkage using AHP factors.

5.2.1.2 Actual FA PANAMAX 07 Performance Assessment Tool

The Operational Level staff was concerned that the utilization of pre-existing JMETs alone might not adequately address the required JMETs or the desired Training Objectives, causing the Operational Level staff to re-examine the measures which would be used for the certification.

As described in section 4.2.4, the Training Objectives devised were eventually matched to either previously existing JMETs or were added to the JMET listing already developed. The additional measures required attributes (i.e. time, percentage, incidents, Yes/No), standards, weights, LDF weights, and skill assignments which were subsequently validated by the staff. The staff had identified the board, center, cell, or bureau that applied, and in some cases which major function, A, P, or E was desired. In all cases where the A, P, E function was not assigned, the researcher made the assignment. The very short period of time available between the time that the researcher was notified of the change in approach and the commencement of the exercise (approximately 48 hours) did require the staff to focus on the attributes, weights, skill assignment areas, LDF weights, and A,P,E choices.

The tremendous variability in experience of the observers, as well as the late delivery of the Observer sheets, precluded gathering data on four successive days as initially planned. The realization that some of the measures applied but could not be adequately assessed because the processes which required observation were completed, was one significant lesson learned by the staff regarding the timing of devising measures for exercise observation. There were 18 measures which were declared Not Applicable as a result of late notification.

There were some measures that simply were not observed, which accounted for another 14 measures that were not assessed. The overall number of measures observed comprised 95.1% of the set of 404 proposed by the staff. This completion rate compares very favorably with U.S. Atlantic Fleet Strike Groups, which typically are able to observe nearly 95% of all possible certification Navy Mission Essential Tasks. Since the ability to collect data for each day to provide trend data was not possible, the goal was established to collect the maximum number of measures at least once throughout the course of the exercise.

The collection of the maximum number of measures, which resulted in 371 measures being observed, still permitted using the performance assessment tool. The collection also permitted comparison of the results between the values for a final day of an exercise, the total T-rating distributions, and the ability to use the LDF methods which had been developed in the FA PANAMAX trial efforts.

The relative inexperience of the exercise observers was of slight concern in that there was the potential for the assessment team to either over estimate the Staff's performance or be overly critical. The objectivity and even handedness of the assessment team contributed, ultimately, to a very realistic and accurate appraisal of the Staff's performance during the exercise. The assessors also provided two other vital functions for both this research effort and for the long term benefit of the Operational Level staff.

First, the assessors were able to validate the measures and standards provided. One of the agreements required to assure timely delivery of all the assessment products was that the assessors would be the judges of the applicability of any measure attributes and standards devised. The assessors were instructed to view the attributes and standards critically, and if changes were necessary, that their input would be the authoritative word. A total of 11 changes were requested, these changes consisted of converting seven quantified attributes to Yes/No values, and changing the T-rating ranges for four other measures. Additionally, there were four JMETs added to the initial set, with the appropriate attributes and standards included. The changes described still resulted in an overall accuracy of the proposed standards of 96.3%. The importance of arriving at an

agreed to set of measures cannot be over stated since this set would also be the set used to report readiness to the COCOM as well as in the Defense Readiness Reporting System (DRRS) for any other Service or Higher Headquarters staff to view.

The second value added by the staff was to complete the JMET assignments for all the Training Observation descriptions. The researcher was able to assign 278/414 (67.1%) quite easily, but was not comfortable assuming which JMETs would be most appropriate for the remaining Training Observations. Using the provided "Rosetta Stone" spreadsheet, the Operational Level staff was able to complete the assignments in about half a day. This not only assured ease of reporting within DRRS, but also demonstrated the utility of the "Rosetta Stone" approach.

Although multiple days of observation would have been the ideal situation, the widely variable experience of the staff generally precluded such an effort. There were isolated instances of multiple observations, such as those done by the Joint Personnel Recovery Cell (JPRC) that provided the opportunity to view the "learning curve" experienced by this portion of the staff. That set of observations is shown in graphical format in Figure 18.

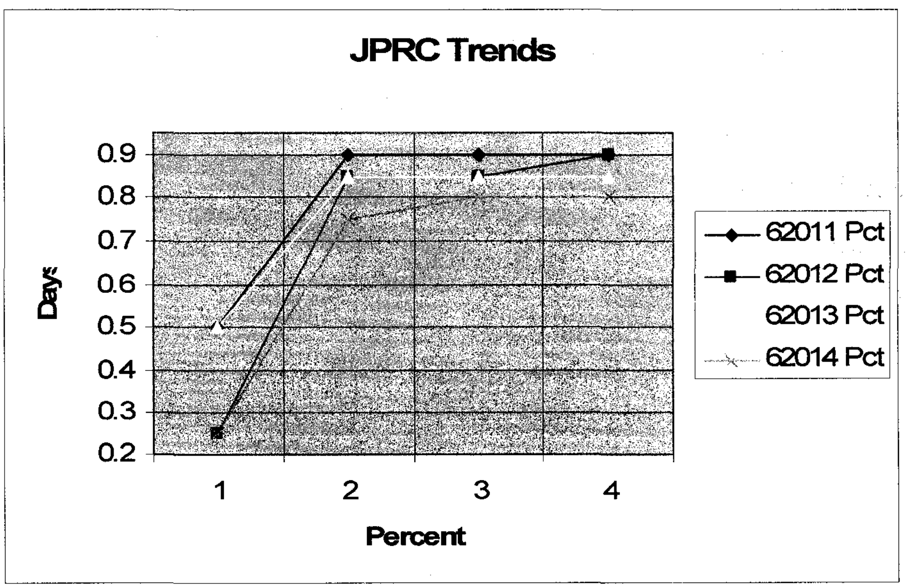

Figure 18: Joint Personnel Recovery Cell Performance Trends

It was interesting to the researcher that the performance trends appeared to mirror the general trends which were observed for the COPS area (predominantly Execute type actions) as depicted in Figures 12 and 13. The applicable values for the graph are provided in Table 14:

	Day 1	Day 2	Day 3	Day 4
62011	0.500	0.900	0.900	0.900
62012	0.250	0.850	0.850	0.900
62013	0.500	0.850	0.850	0.850
62014	0.250	0.750	0.800	0.800

Table 14: PRC Trend Value Data[30]

The overall data collection effort did provide the Commander with a numerical assessment of the Staff's performance at the end of the exercise. This numerical assessment also mirrored the verbal comments from the Observer Trainer (OT) staff from Joint Forces Command. It might be argued that the inexperience of the staff might lead to the complete adoption of observed values from the OT staff, and thus, the similarity would be fait accompli. An important consideration in this regard is the fact that the OT staff does not use a numerical evaluation method as its basis for assessment. Further, although the JMETs required for COCOM are well known by the OT staff, these are not specifically addressed in isolation. The operational level staff is viewed in both an OE view and APE view, and there is an expectation that the ability to link performance from the OE level to the APE needs of the Commander has to exist to permit the Commander

---

[30] 62011 – Integrate joint, multinational, and interagency capabilities and coordinate Personnel Recovery operations, 62012 – Monitor component recovery capabilities. 62013 – Maintain 24/7 watch team, 62014 – Ensure timely dissemination of lessons learned and or time sensitive PR data

to function as required, but there is no mathematical formula in use to view this relationship.

The preceding discussion is particularly noteworthy since the operational level staff in question was the first to attempt a numerical assessment of performance. Although the agreement between the JFCOM OT staff and the operational level assessor evaluation was obviously the desired end state, the concurrence was not known until the final After Action Review (AAR) on the day subsequent to exercise completion. No pre-brief was available to the researcher or the Operational Level training staff to assure concurrence.

The overall staff performance in terms of OE and APE is portrayed in Figures 19 and 20 which follow. The high scores for the Future Operations (FOPS) and Intelligence (INTEL) groups as well as the lower scores for Future Plans (FP) and Command Element (CMDE) are noteworthy. The Commander had stated three desires as major objectives of the Commander for the Joint Task Force Capable Headquarters certification effort, and especially for the exercise. The first was the ability to view his Staff's performance and requirements from a doctrinal standpoint, which was accomplished by the assignment of every Training Objective to a corresponding JMET measure. The second was the development of "kneeboards"[31] so that the JMETs could be accomplished proficiently by any follow-on staff members without a great deal of experience (72). The final objective was the ability to numerically assess the Staff's performance to validate the effectiveness of the training received before the exercise and the utility of the kneeboards utilized during the exercise. The numerical assessment would also serve as the basis for reporting to higher headquarters. The additional benefit of the numerical assessment was that there would be a baseline value for an operationally ready (T2) staff so that the level of

---

[31] Kneeboards are the laminated procedure cards carried by military aviators which are used to carry out almost every task in an aircraft. These procedural guides contain only the key steps necessary to complete a specified task, and thus are typically a single page, printed in large fonts and can be fastened to the aviator's thigh just above the knee for very quick reference without using the hands.

achievement required for future exercises would be known. There would also be a potential method of judging the decline in performance due to staff rotations or a lack of proficiency training since each staff member was assigned to specific JMETs by virtue of his or her OE function.

The values for both FOPS and INTEL were reflected in the comments of the JFCOM OT members. These values could also be directly attributed to the very well formatted kneeboards utilized by both groups. The kneeboards were so well devised that they were also specifically addressed in the AAR by the OTs. The lower than expected values for FP and even the CMDE were also discussed at the AAR. It would be unfair to characterize the low overall score depicted in Figure 19 in terms of performance alone.

Day 4 Observations	
CMDE :	83.4%
COPS :	81.8%
FOPS :	89.0%
FP :	69.9%
INTEL:	87.0%
IO :	82.0%
LOG :	91.1%

Figure 19: Operational Level Staff Observed Performance Values (OE)

Day 4 Observations

Assess

- Composite APE Result: 86.2%
- AHP Linked Result: 85.2%

Plan

- Composite APE Result: 82.9%
- AHP Linked Result: 83.7%

Execute

- Composite APE Result: 84.7%

AHP Linked Result: 83.3%

Figure 20: Operational Level Staff Observed Performance Values (APE)

Certainly some aspects of performance resulted in the lower value, but several other factors contributed as well. The FP area had far fewer observations than the other OE areas (only 13 of 371, approximately 3.5%). The next lowest areas (INTEL, IO, and LOG) had either 32, 31, or 29 measures respectively, still approximately 8.5% of the total but more than double the FP area. Additionally, the nature of the FP area is such that most observations should be conducted well before an event occurs to most effectively and completely assess that particular OE.

One consideration for the lack of observations was that many of the proposed assessment criteria were not able to be observed, but of the 32 measures which were either Not Applicable or Not Observed, only one applied to the FP area. The low numbers of measures was addressed as a factor in the overall assessment while debriefing the Commander. Doubling the number of measures assessed might not have improved the performance, but the impact of notably deficient measures might not have been as significant in the overall assessment. There was no explanation for the low numbers of observable Training Objectives other than a lack of consideration for a larger number of

training objectives and the lack of foresight in assessing the FP area earlier which would have been more appropriate.

Even with those factors being considered, there were some performance shortfalls in the FP area, and these were also similar to those observed in the Strategic Communications (SC) bureau of the CMDE group. The primary issue for both was cited as the lack of a procedural methodology to follow when events became especially dynamic. The lack of kneeboards did not permit a methodical approach to unanticipated eventualities, which caused a more fragmented response than would be desirable. The OT team provided multiple examples of previous kneeboards that could be adopted to assist either OE in achieving the desired end state.

The positive and less enthusiastic comments from the OT team confirmed the beliefs of the Commander regarding the underlying reasons for the Operational Level staff performance and reinforced his long standing belief that kneeboards were a key element in repeatable and reliable process execution.

The viability of the LDF method, whether as a direct measure of APE or as a means of linking OE performance to APE performance as a confirmation of the AHP linkage was not well established during the examination of results from the original FA PANAMAX performance assessment tool. The LDF values obtained for the final observation values were an improvement over previous results, but they still left room for doubt as to the utility of this method regardless of the desirability and elegance. The values for LDF for all factors, only the high or medium value based factors (Limited LDF), or as the result of a second level aggregation are depicted in Table 15 below:

	OE Value AHP Linked	APE Value	LDF Value	Limited LDF Value	Second Level Aggregated Value
Assess	0.8530	0.8621	0.8054	0.9009	0.7325
Plan	0.8379	0.8289	0.7570	0,8281	0.4069
Execute	0.8336	0,8474	0.8165	0.8715	0.5934

Table 15: LDF Values from the Operational Level Exercise

The values for all areas were in much better agreement than any of the previous results obtained from the LDF method, and did indicate the stronger or weaker areas appropriately. The LIMITED LDF approach over estimated the values obtained from either the AHP linked method or the APE calculations in two of three instances, but even in these cases, only one value (the Assess parameter, exceeded the 3% agreement desired, but only by 0.8%). The LDF method values were slightly lower than those obtained from either averaging method, but still were not more than 8% different, which could still be useful. The second level aggregation values were also improved from the results observed from the trial performance assessment tool, but still didn't appear to merit consideration for routine use since the values were up to more than 40% different (Plan) from the averaged or linked APE values.

The issue of what could be done, if anything, to permit the LDF method results to have the utility that the averaging methods appeared to provide was the only aspect of the performance assessment methodology which did not prove to be as successful as had been hoped for during this research.

5.2.2 Cost Modeling Results

The strategy for modeling costs for the Operational Level staff provided the chance to expand upon the methodology used for both the MCO model and the Noble Resolve 07-2 model. Since there was also involvement with JFCOM, there were costs which could that could be captured from that command as well as the costs incurred by the Operational Level staff. A significant difference from the other two use cases which will be discussed later in this dissertation existed in that there were costs that could be identified in all four major cost categories. The Commander expressed a very keen interest in identifying all the possible costs, especially those which would be either indirect or one-time/event specific, so that these could be examined and a determination made as to whether these particular costs could be avoided, converted into a routine cost, or reduced for future events. The Commander was also interested in knowing what resources had been expended that did not contribute to the certification effort and the reasons for those expenditures.

The very unique situation for this specific operational level staff may not be directly applicable to all similar staffs, but there are elements of the certification effort that can be applied to other staffs which could avoid or reduce costs, which would be an immediate ROI from this initial certification effort. Demonstrating ROI for a single event is not as dramatic as sequential events because the only ROI available is that which can be drawn from comparing the relative expenditures for specific JMETs for the readiness received, however more potential ROI opportunities will be discussed in the subsequent section.

One of the important elements that could be shown for the resources expended is the ratio of costs for specific JMETs and a review of the importance of these to the overall certification effort. The core JMET listing was provided in section 5.2.1.1. There

were a number of other JMETs that were necessary to support the certification effort, and a careful review of the applicability of costs which might be attributed to any of these JMETs was required from the Operational Level staff.

A procedure developed in the Major Combat Operations use case for instances where costs could not be attributed to a single JMET but had to be split among several JMETs was applied to the FA PANAMAX cost model. The Operational Level staff was consulted on the splits and the data shown reflects the apportionment of resources to JMETs arrived at by the staff. The cost categorization and identification by JMET was easily accomplished for most of the cost data submitted. Those that were not readily identifiable were categorized (in the Personnel area only) as "across-the-board" (ACB) meaning that any cost figure was evenly divided among the seven core JMETs. The overall costs attributed to the certification JMETs are shown in Table 16 as well as in Appendix J.

Category	OP	Cost	Category	OP	Cost
Direct	5.5	$108,119.	Indirect	5.5	$77,000.
Personnel	1.1	$269,690.	One-Time/Event Specific	1.1	$71
	3.1	$269,690.		3.1	$71
	4	$316,147.		4	$71
	5.2	$269,690.		5.2	$22,010.
	5.5	$311,068.		5.5	$27,466.
	6.2	$269,690.		5.7	$71
				6.2	$71

Table 16: FA PANAMAX07 JMET Cost Allocations

Table 17 shows the ratio of expenditures per JMET.

JMET	% Allocation
1.1	12.1%
3.1	12.1%
4	14.25
5.2	13.1%
5.5	23.6%
5.7	12.7%
6.2	12.1%

Table 17: FA PANAMAX 07 JMET Resource Allocation by Percentage

5.2.3 ROI Results

The only significant drawback encountered with the performance and cost modeling efforts revolving around the Operational Level Staff certification was the fact that ability to rigorously capture ROI does not yet exist. There are however; potential ROI targets which can be forecast and those are worth discussing.

The baseline JMETs used in the certification do represent the core set of missions which can be realistically expected of this Operational Level staff by the COCOM. As shown in section 5.2.1.2, the level of performance (which was deemed sufficient for certification) achieved sets a standard that can be held up as the benchmark for future certification and training events. The development of the JMET standards provided a minimum set of JMETs which can be the basis for proficiency training and the development of training scenarios for newly reporting personnel as well as for any future certification event.

The very detailed cost information which the Commander directed be collected can be put to immediate use for several purposes, not all of which would be specifically related to ROI, but which would easily translate into ROI after any subsequent event.

Since this was the first Operational Level JTF Capable Headquarters certification event conducted for any Service, a baseline for anticipated cost for such an event did not exist. However, the value of archiving ROI cost data for specific JMETs (which could prove to be common with other staffs) would be available for use to any follow-on JTF capable headquarters certification efforts, and could also assist refinement of costs for sustainment efforts. The indirect and one-time/event specific costs also provide an area which can be examined and potentially would result in cost avoidance. From a larger service or operational level perspective, direct comparison of these costs between two events, especially in terms of similar JMET baselines provides a measure of ROI for these events which has not been previously available.

The length of time required to reach sufficiency, the areas that proved to be the most challenging, and the experience of the staff are all firmly established and can be utilized in several ways to yield differing views of ROI. The most obvious ROI opportunities for this specific Operational Level staff will occur after a subsequent event when the costs for the certification can be directly compared to those from the 2007 event. The changes (reduction or elimination of either a part of or significant amounts of the indirect and one-time/event specific costs) in costs will be a direct measure of ROI for a common and performance assessment based certification.

If further reductions in costs can be achieved by reducing the time the exercise must be conducted to achieve acceptable performance, re-use of the scenario, or potentially conducting the exercise in concert with another operational level commander exercise, other measures of ROI will be available. The ability to reduce travel by either reducing the number of personnel required to be distant locations or through more widespread use of tools such as video teleconferencing or improved situational awareness capabilities is one of the greatest ROI potentials for such an exercise.

One final method, and this listing is not meant to be exhaustive or complete, merely illustrative of possibilities, would be to tailor the preparation training to reduce scenario time or scope, thus reducing contractor support, or to review trends from the 2007 event as well as the estimate of Staff readiness at exercise start to limit the number of external observers and specialized support personnel required, which would reduce costs

5.3 MCO Model Results

5.3.1 Performance Assessment Results

Several changes to the performance assessment tool were mandated as a result of the initial instantiation of the model, and these were incorporated into the MCO use case (as well as the other two use cases discussed in this dissertation). The JMETs used to formulate the performance assessment tool were only those which had a value as discussed in section 4.4.3. The other changes which were incorporated were the reduction of redundant measures which was discussed in section 4.4.4 so that a measure's value was entered only once in the applicable OE area and once in the corresponding A, P, E area. Also as noted in section 4.4.3, the original version of the MCO performance assessment tool included 1225 measures, but that version also had a significant number of redundant measures. When all of the changes discussed above were incorporated, the revised version had a total of 692 measures. This would still require approximately 25-35 observers, but inevitably some measures would either be declared Not Applicable or would circumstances would prevent some observations from occurring. This performance assessment tool was submitted to both TTGL and FFC for review and validation.

The review of the skill assignment areas and the numbers of measures that applied as well as a review of the OE measure totals was conducted for this formulation of the performance assessment tool as well. There were some OE categories which had low numbers of measures, but still enough measures that could impact the overall outcome, the fewest being 43 measures for the IO area. The maximum number of measures was 194 for the COPS area. Similarly, there were some skill assignment areas that had lower

numbers of occurrence than the others, but the reassignment of skill assignment areas discussed in a subsequent section of this dissertation resulted in a more balanced distribution of skill assignments than had been observed during the initial instantiation of the performance assessment tool. These reductions were submitted to both TTGL and FFC. Neither organization believed that there should be further combinations or reductions, although it might be possible to make a further reduction of skill assignment areas in any future work with sufficient discussion with the Commander and staff involved.

The lack of historical CFMCC performance data posed the same challenges in model assessment as with the initial performance assessment model testing. The concurrence of averaging methods and the agreement of either LDF method with the averaging methods was the only way to assess if the formulation was properly calculating the observed values being used. The distribution of observed values used for this use case, depicted in Table 18 was nearly the same as those for the initial performance assessment tool depicted in Table 11:

T2(%)	T3(%)	T1(%)	T4(%)	Overall (Predicted)
67.3	19.8	6.7	6.2	.7795

Table 18: MCO Model T-Rating Percentages

This distribution is similar to the normal distribution of results discussed in section 4.4.5 in Table 10 for a Day 2 T-rating distribution. The results for the MCO model for OE performance ranged from 0.798 to 0.823 which was better than the agreement seen for the initial model, but still differed from an almost normal distribution of T-rating observations. The two alternate formulations for T-rating observations

depicted in Table 11 (.8025) and Table 12 (.8075) respectively, were much closer to the results of the OE areas.

The results for the APE areas as well as the aggregation of values from the OE values were not as consistent as those observed for the OE values alone. The values calculated for APE ranged from 0.754 to 0.817 which are not greatly different from any of the predictions but the range of values was greater than previously observed. Notably, the Assess value was much lower than either the Plan or Execute values. This was true regardless of which AHP method (either importance or number of measures) was applied to the OE areas to compare the OE results to the APE results as well. In these cases the range of APE values ranged from 0.789 to 0.792 (importance) or from 0.793 to 0.795 (number of measures). None of these differences was more than 3.5% different (comparing APE to the aggregated OE value) but the difference was not seen in the initial model and was noteworthy.

The values for LDF reflected the results seen in the Initial model after changing LDF factors to High/Medium/Low. When a large number of skill assignment areas were used to create a LDF value in the OE categories, the values tended to be very small, sometimes as low as 0.381. The Limited LDF values were more consistent with the values calculated for either normalized averaging or floating weight averaging, in all but two cases being with 5% of the either averaging method.

The LDF results for the APE categories more closely approximated the results obtained from either averaging method, and had similar magnitude (no worse than approximately 4% in the Assess instance) to the results noted in section 5.1.1.2 and could be considered as a method for reporting performance assessment. The LDF results were not as accurate, especially in the Assess area, and still could not be considered a viable method for reporting performance. The values observed are shown in Table 19:

	LDF	Lim LDF	Averaging
A	.5647	.7992	.7544
P	.7762	.8048	.8165
E	.7120	.7861	.7927

Table 19: LDF vs. Averaging Values for A, P, E

One method to relate the performance of each of the Organizational Elements to the overall APE process is to use the values from AHP to calculate each Element's contribution to the APE process. Could there be another way to relate the OE performance to APE? A possible solution appears to be achievable by conducting a second aggregation of the OE contribution to APE by viewing each contribution within the APE construct. The method used to calculate the APE performance discussed in all previous discussions is to look at each Organizational Element's Assess, Plan, or Execute measures irrespective of the skill assignments. If each Organizational Element's Assess, Plan, and Execute measures were further segregated by skill assignment, then each OE could be viewed by a second LDF equation which would be a more direct method of relating OE performance to APE. This approach was first attempted with the MCO use case performance assessment model. The results of the second level aggregation are shown in Table 20. While these results are better than those from the LDF calculations in the OE categories, the results still would not be sufficient for use in performance assessment.

	LDF (Aggregated)	Averaging
A	.6328	.7544
P	.5341	.8165
E	.6719	.7927

Table 20: Aggregated LDF vs. Averaging Values for A, P, E

Viewing LDF using a limited number of factors as in the APE case, using only limited factors, or even attempting the second level aggregation LDF methods showed promise. The potential for an LDF method as the most vivid method of showing performance - since a poor performance in a High importance area or substandard performance in multiple Medium importance areas would be immediately seen in the product relation - and the factor causing the lower product remained as an interest throughout this research.

5.3.2 Cost Modeling Results

The first set of cost data that was made available for this research was delivered from Fleet Forces Command. The data afforded to the research effort comprised the costs that had been accounted for during two consecutive years of FST-J events as well as the costs incurred during the July 2006 Joint Task Force Exercise (JTFEX), a LVC event. FFC expressed an interest, in concert with JFCOM's desires for this research, to compare not only LVC data to synthetic data but also to compare successive years of synthetic data. Both views could be insightful and might provide decision makers with a clearer understanding of the most effective use of resources in either type of exercise.

The data supplied for both 2006 exercises was clearly identified by exercise type, so differentiating any virtual or constructive resources utilized in support of JTFEX versus FST-J was not an issue. Similarly, the costs were easily separated into the appropriate costing category.

The cost accounting methodology used at FFC was not able to capture either Indirect or One-Time/Event Specific costs, but this is probably because no specific effort or formal methodology to do so had been identified or requested before this research. That does not imply that neither of those categories would have been applicable, just that no requirement to capture those categories of costs had been declared, so neither category could be accurately populated.

Another issue regarding costs for the JTFEX was the fact that no aviation fuel costs or flying hour costs for U.S. Navy aircraft, commercial aircraft (often used to replicate Opposition Force (OPFOR) aircraft), or U.S. Air Force were made available to FFC. The eventual cost models would be able to show personnel expenses for Commercial aircraft, U.S. Air Force aircraft, and for the range services and environmental impacts of strikes, but this is only a partial story. The more complete picture that the aforementioned costs would provide is evident, but as long as all costs are consistent, knowing the costs are not included still permits a formal method of detailing costs that could easily be modified should the costs be released at a later date.

One final area of debate within FFC regarding costs required resolution for the cost modeling to be consistent. Ships require pier service support while in port to sustain normal operations and to permit repairs. These costs are primarily the water, sewage, electrical power, and steam which are supplied to each ship. There was a view within some portions of FFC that these costs should be included in the synthetic training costs since the ships did receive these services and completed training. The N7 directorate reviewed this approach but noted that whether a ship was manned and participating in a FST event or not, the costs would be incurred. The very specific network connectivity costs, any contractor support, or specialized installation costs were already identified and those were the unique costs for the event. This reasoning led to a decision not to include port service costs in the costs to execute the FST-J events.

As noted earlier in this section, the cost data was segregated by FFC so that the major task was to appropriately assign JMETs to the cost data. The availability of some specialized units such as the Patriot batteries could have permitted the assignment of

costs to a three digit JMET, but the other assets such as the cruisers, destroyers, and AWACS (67)[32] had missions that included but were not exclusively Theater Air Defense. The more general two-digit JMET, OP 6.1 Provide Operational, Air, Space, and Missile Defense, could be applied to all the units involved since their missions also included air defense.

The JMET assignments for Live and Synthetic Direct Costs are shown in Tables 21 and 22 below:

	Live	Applicable JMET
Port Services	$110,608.	OP 1.2
Fuel	$5,791.459.	OP 3.2/6.1

Table 21: Direct Costs with JMET Assignments

	Synthetic	Applicable JMET
Network Costs	$153,439.	Op 3.2/6.1

Table 22: Direct Costs with JMET Assignments

The reason that there is a split in the JMET assignments is that the units involved were not all Air Defense units but OP 3.2, Attack Operational Targets, applied in some

[32] Airborne Warfare and Control System

part to all units assigned. The network costs were split in the same fashion. The costs to get the ships underway (Port Services) are assigned entirely to OP 1.2, Conduct Operational Maneuver and Force Positioning. There is no analogous cost for the synthetic costs because the units are placed in theater and no time in the scenario is used for transiting units or re-positioning prior to commencing the scenario. The costs for fuel account for any maneuver that the units must carry out when the scenario is ongoing. Where multiple JMETs are assigned, the costs are divided according to the applicable mission area. In the instance described above, three of the eleven vessels assigned had no Air Defense role, so the fuel costs were split accordingly. In the synthetic case, only six vessels could be connected, and again one of the six had no Air Defense role, so the costs were divided by six and one-sixth was assigned to OP 3.2.

A similar methodology was followed to account for the Personnel costs in both the Live and Synthetic venues. These costs are also illustrated in the Tables 23 and 24:

Area	Specific Category	Cost	JMET
Training Support		$471,580.	OP 3.2
	Commercial Air Services	$663,305.	OP 6.1
	Environmental	$73,081	OP 5.1
Joint/Coalition	USAF-AWACS	$343,719.	OP 3.2/5.1/6.1
	– Patriot	$351,718.	OP 6.1
		$1,460,592.	Op 6.2
		$1,587,600.	OP 6.2

Table 23: Live Personnel costs with JMET Assignments

Area	Specific Category	Cost	JMET
Training Support	Exercise Labor	$545,205.	OP 3.2/5.1/5.2/6.1
	Surge Support	$940,351.	OP 3.2/5.1/5.2/6.1
	Connectivity	$168,055.	OP 5.1
Joint/Coalition	USAF –AWACS	$34,600.	OP 3.2/5.1/6.1
	USAF-ACS/CRC	$3,775.	OP 5.1/5.2
	USAF – CAOC	$19,000.	OP 5.1/5.2
	– Patriot	$96,722.	OP 6.1
		$31,153.	OP 6.2
		$37,726.	OP 6.2

Table 24: Synthetic Personnel Costs with JMET Assignments [33] [34]

The reasoning for the JMET assignments and the splits followed the same methodology as for direct costs. There is no question that additional JMETs or even a split to three digit JMETs could be achieved for the synthetic and live costs where personnel are concerned. Certainly arguments could be made that additional areas of OP 5 such as Provide Operational Command and Control could be considered. Similarly, there could be additional OP 6 assignments for the Surface Combatants, but the time spent in other areas such as surface defense or anti-submarine warfare for the U.S. Navy cruisers and destroyers was minimal next to the strike and air defense tasking. The reverse is true for the British and German units assigned. In future events, both of these countries as well as a number of others may well assign units with much greater air defense capability; it is therefore noteworthy that costs will need to be split accordingly.

---

[33] ACS – Agile Combat Support (JP 1-02), CRC – Control and Reporting Center (JP 1-02)

[34] CAOC – Combined Air Operations Center (JP 1-02)

The second FST-J event with costing available was conducted in July 2007. Once again, FFC was able to very clearly delineate how the costs could be split. As for the FST event in 2006, no pier services costs were included in the overall costs. Additionally, since both the United Kingdom and Germany had established the capability to connect to the FST event without U.S. assistance, no costs were incurred for their personnel in the 2007 event.

The specificity of costs, especially in terms of the Joint/Coalition costs was much more detailed than in the 2006 event. A complete listing of the costs is shown in Appendix G. A summary of the costs is provided in Table 25:

Area	Specific Category	Cost	JMET
Direct	Networking	$222,514.	OP 3.2/5.1/5.2/6.1/6.2
Personnel **Training Support**	Exercise Labor	$75,250.	Op 3.2/5.1/5.2/6.1
	Surge Support	$208,930.	OP 3.2/5.1/5. 2/6.1
	Joint/Coalition	$149,357.	OP 3.2/5.1/5.2/6.1/6.2

Table 25: Synthetic Costs with JMET Assignment

Even a cursory review of the costs detailed in Table 24 indicates that substantial savings were made over the 2006 FST-J event. Significant infrastructure investments were made between the two FST events which improved the communications capabilities for the facility at Dam Neck, VA. Similar investments were made by both Coalition partners that had two effects. The need for U.S. contractor personnel to assist in both scenario set up and execution was reduced to a single small team. The second effect was that the improved communications permitted the use of liaison officers for both the United Kingdom and Germany who were already stationed in the Norfolk area, so no U.S. funding was required to move a Coalition staff to Norfolk to support the exercise. These factors eliminated almost all the travel associated with FST-J 06-1, which markedly reduced the overall costs for Training Support.

The methodology for assigning costs as delineated above and in the various tables achieves the very important goal of being able to view costs in terms of accomplished training as well as being able to compare them directly due to the JMET assignments.

5.3.3 ROI Results

The cost modeling above permitted the first, although admittedly limited, application of ROI utilizing JMETs. The data forwarded by FFC enabled three very useful views of ROI. Another benefit of the JMET assignments not obvious until all the costs had been categorized was the fact that it is possible to view the relative costs of each JMET against the total costs to get an estimate of the relative value of the training received with regard to the resources expended. An example of the tremendous potential utility of this application will be expanded upon later in this section.

The first useful view of ROI, and one of great interest to not only FFC but to JFCOM and almost all organizations, is the direct comparison of LVC against synthetic costs. When data is available for the 2007 JTFEX event, multiple comparisons will be possible. Since only 2006 JTFEX cost data was available to the researcher, only the 2006 events could be compared. The LVC and synthetic costs are shown, and the relative advantage of synthetic training is noted in the ROI column. A second set of figures shows LVC to synthetic, but in this instance only the JMETs that match are shown. As delineated in Tables 21 – 25 and discussed above, OP 1.2 and OP 5.2 were not used in both events, so the ROI with these costs removed is also depicted in Table 26.

	LVC	Synthetic	ROI
All Costs	$10,853,662.	$2,021,836.	5.37
Only Similar JMET Costs	$10,853,662.	$1,708,782.	6.35

Table 26: LVC vs. Synthetic Costs – ROI

The third view of ROI is the comparison of the synthetic costs in successive years for the FST-J events. Since both events were viewed with the same JMET basis, directly comparing costs was greatly simplified, and the ROI is shown in Table 27:

	FST – J (2006)	FST-J (2007)	ROI
Total Event Cost	$2,046,879. .	$656,051.	3.12

Table 27: FST - J Costs – ROI

5.4 **Noble Resolve Model Results**

5.4.1 Performance Assessment Results

The description of the methodology employed to construct the performance assessment tool for Noble Resolve (NR) 07-2 was discussed in Section 4.4.6. The various changes in weighting methods, LDF assignment, and single assignments of measures to proper OE or APE categories was utilized at the outset for this performance assessment tool. The changes required because of the application to a Homeland Security scenario eliminated 166 measures and made the tool more reasonable in scope than the originally proposed tool. Also as noted, the addition of the Radioactive Dispersal Device (RDD) event caused a slight increase (49 measures) in the tool, but the ability to quickly revise the tool to fit the needs of the scenario and the training audience was a very useful demonstration of the stability of the performance assessment tool.

Since there had been no pre-agreement with either the J-9 staff (who were executing the coordination center at JFCOM), or other organizations (principally the states and government agencies involved in NR 07-2), assessment was done to capture major JMETs accomplished. The assessment of some performance was done to view a

typical second day of a well developed exercise using observations from the Joint Operations Center voice, data, and video feeds as well as from the results of the daily After Action Reviews (AAR). The researcher submitted his observations to J-9 experiment supervisory personnel for concurrence. The same process was followed for day 4 of the exercise.

The performance results, based on AAR input and observation varied somewhat from what might have been predicted in that the resolution of very low performance on the first day of the experiment was dramatic. There were very few completely unsatisfactory observations (eight on day one, three on day three), but nearly 30% (165/591) of the results on day one were in the T3 (Below Average) category. This number was reduced by almost exactly half by the last day (83/591). Similarly, there were a lower number of T1 (Above Average) observations on day one (26/591) than might be expected from a trained operational level staff, but the number of T1 ratings increased by almost 300% (76/591) by the last day. This result could be attributed to the fact that the Emergency Operations Center (EOC) watch standers in most State EOCs are not professional watch standers by either job description or background. Though they are periodically trained, they do not receive the same level of training as their military staff counterparts. Nonetheless, by the last day of the experiment, the teams were far more proficient in the crispness of verbal reports, and information flowed much more easily. There was also a much better situational awareness evident in written reports, the master common operational picture, and of the information requirements of the other agencies involved. The changes in T-ratings are shown in Table 28:

	T1	T2	T3	T4
Day 1	26/591	392/591	165/591	8/591
Day 3	76/591	429/591	83/591	3/591

Table 28: Noble Resolve T-rating Evolution

Unlike the Major Combat Operations use case, which relied almost entirely on the researcher's experience, or was a combination of multiple people's experience with Major Combat Operations, Noble Resolve was a focused experiment with the Defense Support of Civil Authority (DSCA) (2) and Homeland Security (HLS) issues. This required a more cooperative approach in determining the applicable JMETs. The addition of the Radioactive Dispersal Device (RDD) scenario made the need to collaborate in devising appropriate measures more critical.

The importance of the collaborative approach manifests itself in the use of the Analytical Hierarchy Processing (AHP) methodology (40). In the two previous performance assessment constructs, the JMET driven approach was appealing due to the lack of personnel to survey, or the very small numbers available (there were only 10-20 personnel available on an irregular basis serving in the CFMCC role) to poll in order to assess the relative importance of each OE area in relation to the APE functions. For Noble Resolve, there were certainly a larger number of people available, but they were in each of the state or territory EOCs, so not available to poll. The collaborative effort in determining the applicable JMETs provided an alternate means of judging the importance of each OE within the APE functional areas, in a JMET driven fashion.

The collection of data, even if only done via communications and AARs provided an opportunity to evaluate the performance tool's ease of use. The loss of the ability to utilize a relational database late in the research effort was somewhat disruptive, and if multiple events required archiving would be a necessity. Fortunately, in the case of a single experiment, multiple days can be easily replicated using EXCEL™ tool sets. The total time required after experiment scenario completion each day did not exceed two hours. This was the first time anything other than hypothetical data was available for assessment, so the results were of great interest.

Tables 29 and 30 show the OE assessed areas and the values achieved using AHP to estimate the APE values. The second APE column shows the values obtained by directly calculating the observed results for each major functional category. There are values for day one and day three.

		Day 1			
CMDE	.8453	A	.7812	A	.7901
COPS	.8007	P	.7923	P	.8007
FOPS	.8008	E	.8255	E	.8285
FP	.8158				
INTEL	.8098				
IO	.8360				
LOG	.8362				

Table 29: Day 1 Performance Results

		Day 3			
CMDE	.8644	A	.8420	A	.8457
COPS	.8247	P	.8462	P	.8397
FOPS	.8327	E	.8396	E	.8425
FP	.8365				
INTEL	.8343				
IO	.8751				
LOG	.8576				

Table 30: Day 3 Performance Results

None of the prediction models discussed in section 4.4.5 matches the performance observed during the experiment, but the expectation that overall T2 capability could be at nearly 70% on day 2 of an experiment or exercise appears to be reasonable. The staff's or team's ability to eliminate unsatisfactory performance and minimize below average

performance also is evident, but does not follow a normal distribution (i.e. both above average and unsatisfactory performance trend together), rather, above average performance, once attained tends to be sustained.

Although it is not entirely appropriate to merely average results from the OE category to project the APE results, it would be quite normal for an observer to speculate as to why the values obtained for the OE do not appear to "average" to the APE results. The result of averaging these values is shown in Table 31:

	OE Average	APE Average
Day 1	0.8207	0.8064
Day 3	0.8465	0.8426

Table 31: Average OE and APE Values for NR 07-2

Clearly the averages for Day 1 are not equal, and though it would appear that the averages for Day 3 are nearly equal, that result is more coincidence than not. The fact is that the performance of the staff in the APE realm is an integrated effort based upon the interactions of the various OE inputs. These inputs are not necessarily linear, but the use of the AHP method provides a linear estimate. If the desire was to view a more representative "overall" value for performance, then there would have to be a relationship developed that accounted for the relative contribution which the Assess, Plan, and Execute functions have to an overall outcome. Earlier in this dissertation the researcher noted that the predominant functions of any Operational Level staff were Planning and Assessment. Using this approach, a simplified value for performance would allow a more direct comparison of the OE averaged value and the APE results, but as shown in Table 32, the values could provide nothing more than an estimate of overall performance. The value of presenting a single "score" for performance did not appear as important at this point in the research as the ability to show how each Organizational Element performed and how these performances related to the APE performance.

	OE[1]	APE[2]
Day 1	0.8207	0.8008
Day 2	0.8465	.0.8415
1 – Strictly numerical average of OE values		
2 – Based upon a .25 (A), .65 (P) .10 (E) contribution from Tables 28 and 29		

Table 32: Comparison of OE and APE Values for Performance

As noted in the previous paragraphs, the values for either day appear to be comparable, especially for Day 3. Since all OE values exceeded 0.80, it is reasonable that the overall values for APE would do so as well. That may be sufficient in a very broad overview, which is sometimes all a Commander might desire. The use of such estimates as the basis for budgeting or resource allocation decisions would not represent the rigor required to arrive at the OE or APE results, and thus would seem to be no better than using stop light charts.

Another unexpected opportunity arose in October 2007 when Top Official (TOPOFF) 04 was conducted. This exercise had LVC components which had been executed completely synthetically in Noble Resolve 07-2. The biggest LVC contribution was the replay of the RDD event in the Territory of Guam. There were LVC components in the remainder of the scenario, but the RDD event appeared to offer the best potential to capture both performance measures and costs.

The JMETs devised for NR 07-2 for the RDD event were forwarded to Guam in both spreadsheet assessment tool format and Observer Grade sheet format. There was no expectation that the observers input data into the spreadsheets or that the observers on Guam concern themselves with the overall impact that the RDD event had on TOPOFF

04, the sheets were merely forwarded so that the methodology used to show the RDD results for NR 07-2 would be evident and could be compared with the TOPOFF 04 results if desired.

The Observer Grade sheets permitted the observers the opportunity to enter observed values on either a numerical basis or in terms of T-rating, which could be converted to a number by the researcher. Additionally, since TOPOFF 04 utilized the same scenario, architecture, and agencies as NR 07-2, the opportunity to assess (in the same manner as NR 07-2) the performance of the staffs could potentially show the impact of training and provide a contrast of LVC vs. synthetic costing.

The observers for TOPOFF 04 were a completely different group of people, as previously referenced. The RDD event was observed by the personnel on Guam, the overall exercise was observed at the National Operations Center (NOC) in Washington, D.C. by a member of the J-9 Directorate who had coordinated the NR 07-2 experiment in August. The fact that a completely different set of observers were being utilized was also an opportunity to evaluate the usability of the performance assessment tool and observer grade sheets.

The overall results of the RDD event on Guam in terms of JMETs used, JMETs which applied, and JMETs which were altered are shown in Table 33 below:

JMETS Proposed	JMETS Used	JMETS Altered
49	47	0

Table 33: RDD JMET useage for TOPOFF 04

The results of the RDD event during TOPOFF 04, especially in contrast to the results from NR 07-2 were of great interest. The results for both events provided a chance to contrast the survey method for determining AHP coefficients as opposed to

using JMET driven AHP coefficients. An examination of the JMETs chosen for both events brought an issue to light that had not been considered earlier. What if there were no JMETs observed in an APE functional area? How could the AHP method be used in such an instance? The researcher considered the value 1 as the factor, but the very low numbers of factors in each area (no more than 31 and no less than 7) caused the contributions to be less than effective since the total for all factors does need to either equal 1 or at least approach it very closely (40). The solution to this problem appeared to be to place a non-zero value in place of the zero number of factors. Two cases were experimented with, and a value of .01 yielded the most consistent results. There will be more discussion of this point in Chapter 6. The results for the RDD event for NR 07-2 and TOPOFF 04 are shown in Tables 34 and 35:

OE		APE(AHP)		APE	
CMDE	0.8455	Assess	0.8341	Assess	0.8655
COPS	0.8256	Plan	0.8393	Plan	0.8532
Log	0.8508	Execute	0.8498	Execute	0.8010

Table 34: NR 07-2 RDD Performance Assessment Values

OE		APE(AHP)		APE	
CMDE	0.9057	Assess	0.9145	Assess	0.9500
COPS	0.8923	Plan	0.9034	Plan	0.8970
LOG	0.9209	Execute	0.9181	Execute	0.9320

Table 35: TOPOFF 04 RDD Performance Assessment Values

The values for APE either when AHP values were used to link the OE values to the APE performance or as an aggregation of the A, P, E measures do have some

disparity, but with the low numbers of measures available for either the Assess or Execute functions as devised for this experiment the results still represent the contributions of the OE values realistically. Only one set of RDD data was obtained since it was a single day event and no allowance for assessing the planning portion (which may have permitted more JMET measures to be utilized) of the event was made. This was a second instance where low numbers of measures had been evaluated or the use of AHP was impacted by low numbers. The Operational Level Certification had similar issues which were detailed earlier in this chapter.

5.4.2 Cost Modeling Results

The core JMET set used for modeling costs for Noble Resolve were similar to those which would be used for the Operational Level Certification with some notable exceptions. There was no need to view the ability to strike targets, and as discussed in the formulation, the lack of strike requirements also obviated the need for the majority of the intelligence related JMETs. The Operational Movement required for a disaster scenario is more similar to OP 1.1 (Conduct Operational Movement) than OP 1.2 (Conduct Operational Maneuver and Force Positioning (1). There are aspects of Operational Force Protection in a disaster scenario, but the need to protect forces from either insurgent action or a large scale adversary does not apply. There is also some need to protect communications, but not as would be required against an adversary attempting to use information to conduct defensive or counter-offensive operations, so there did not appear to be a need to consider any of the OP 6.1 (Provide Operational Air, Space, and Missile Defense), OP 6.2 (Provide for Operational Forces, Means, and Noncombatants), OP 6.3 (Protect Systems and Capabilities in the Joint Operations Area), or OP 6.5 (Provide Security for Operational Forces and Means) (1) JMETs. There could be a basis for some of the OP 6.5 measures, but the scenario design and areas of concern for the experiment did not readily support assessing that JMET specifically. The JMETs that were most appropriate for cost application were:

OP 1.1 Conduct Operational Movement

OP 4 Provide Operational Logistics and Personnel Support

OP 5.2 Assess Operational Situation

OP 5.5 Establish, Organize, and Operate a Joint Force Headquarters

OP 5.7 Coordinate and Integrate Joint/Multinational and Interagency Support

OP 7.3 Coordinate Passive CBRNE Defense in the Joint Operations Area

OP 7.4 Coordinate Consequence Management (CM) in Joint Operations Areas (JOAs) (1)

The cost assignment process for these costs followed the discussion in section 4.6.4 regarding the Major Combat Operations (MCO) model. In this instance, the only cost category captured was the Personnel category, which is a change from the MCO model and also different from the Operational Commander cost modeling presented in Section 5.1.2. A complete display of the cost break outs is available in Appendix G.

The JMET assignment process also followed the template discussed in section 4.6.4. As observed in the case of the MCO model, some costs could be exclusively assigned to a single JMET, but in others, there was a necessity to ratio the costs in terms of scenario emphasis.

5.4.3 ROI Results

The results are shown in detail in Appendix G, but the costs that would be most interesting to compare if a LVC to synthetic comparison could be undertaken at a future date would be those required for both OP 5 and OP 7. These overall OP groups contained the areas of interest for Noble Resolve, especially in terms of the overarching goals discussed with Section 4.7.1. The ability to exercise command and control (OP 5) and to respond to CBRNE events (OP7) were the major resource allocations for NR 07-2.

Do LVC events result in resource allocations similar to those projected for NR 07-2? If not, why not, and if so, is there any way to reduce the costs or use the resources more effectively? The threat posed by a CBRNE event such as described by Flynn (73) or due to a RDD has caused many localities to develop and practice scenarios which test

all levels of command and control, emergency response, and Federal, State, and Non-Government agency assistance. Is the cost for LVC exercises preparing for these threats in line with the costs projected for NR 07-2? These should be the first areas examined if a LVC cost estimate is obtained. The next best method to ensure proper resource allocation would be to formally assign costs during the next experiment and compare those allocations to those for NR 07-2.

5.5 Performance Assessment Lessons from the Initial Effort

The initial instantiation of the performance assessment model comprised all the JMETs that had been assigned values from the three organizations that had done significant work as discussed in section 4.3. All the Organizational Elements identified within the MHQ w/MOC directive were also used (68). The 28 possible skill assignment categories discussed in Section 4.3.2 were retained as well. The LDF factors were used as described in section 4.3.4, and the original weighting system, which was predominantly populated with .8 and .9 weighting factors, was also retained. In order to evaluate the ability of the model to arrive at consistent results, observed value results were based upon a projected distribution of observed values similar to that which would be seen (as discussed in section 4.4.5) on the second day of an exercise.

The actual percentages of T-rating values are displayed in Table 36:

T2(%)	T3(%)	T1(%)	T4(%)	Overall (Predicted)
68	20	6.5	5.5	.7898

Table 36: Initial Model T-Rating Percentages

This range of T-Ratings in combination with the measure weights and LDF weights should result in scores near the value shown in the overall column of Table 10, but the actual values for each of the OE categories ranged from .808 to .859. When AHP

values representing an estimation of the importance of the areas to each other as related to each of the categories Assess, Plan, and Execute[35] (40) were assigned, the values for the APE categories ranged from .831 to .834. In either case, these values are greater than what would have been predicted by the combination of values as shown in Table 10.

The one area that seemed as though the choices of factors was not appropriate was the LDF product result noted for all views, whether OE or APE. Since the ultimate value for the LDF result would always be a product, it seemed as though the logical manner to proceed would be as discussed in section 4.3.4. The results of this approach, with LDF factors ranging from .8 to .95 typically yielded values for LDF that were significantly different than either averaging approach as detailed in Table 37:

	LDF	Lim LDF	Norm Avg.	Fl Wt. Avg.
A	.5998	.7212	.8290	.8303
P	.5660	.7144	.8284	.8293
E	.5629	.7168	.8250	.8244

Table 37: Initial Model LDF vs. Averaging Results

The Lim LDF refers to looking at a limited number of factors versus all the possible factors. This was done to view the impact of using only the factors that represented the largest number of skill assignment values. Results in the OE area had shown that a large number of factors, typically more than 10 to 15 would result in a very low value of LDF, so one approach was to view only the factors that represented the skill assignments with the largest number of values since this was one method of viewing importance.

[35] Using the AHP conventions of assigning values 1, 3, 5, 7 for importance where 1 is equally important, 3 is somewhat more important, 5 is more important, and 7 is much more important is a well accepted approach to using AHP methods. For this example, the researcher did the assignments whereas a survey would normally be done to ascertain the importance of each area.

After the discussion with Dr. Startin, a revisit of this initial model using the H/M/L methodology for both weighting and LDF was performed (69). The skill assignments were not changed. There was no variation in the Normalized Averaging results in any of the tables, as would be expected. Some small differences appeared in the Floating Weight Averaging results from the original model, but these were also not significant. The major differences appeared in the LDF and LIM LDF values, lending credence to the use of the H/M/L approach for LDF when viewing skill assignment areas as well. The results are shown below and reflect the same APE areas as shown in Table 38.

	LDF	Lim LDF	Norm Avg.	Fl. Wt. Avg.
A	.7059	.7833	.8290	.8283
P	.6722	.7632	.8284	.8259
E	.6937	.8246	.8250	.8239

Table 38: Modification of Initial Model LDF and Averaging Results

The use of the Averaging methods to check against each other and as the reference for the LDF results was entirely because no historical data for this type of approach existed and no exercise data had been collected when these models were constructed. There were no selections of the H/M/L values for either the weighting factors or the LDF factors from any training audience, so the selections were based upon the researcher's experience in the case of the weighting factors or the total number of skill assignments or values in the case of the LDF. The approach was consistently applied, so it appeared reasonable, especially to verify the ability of the models to correctly calculate all needed quantities.

Two other issues in the construction of the models appeared to require a change for any later constructs, the skill assignments and the OE assignments. Though the MHQ w/MOC instruction specifically cites both the Communication Information Systems

(CIS) and Liaison (LIA) areas as distinct parts of any Organization (68), these OE areas did not have a significant number of measures in relation to the other areas. Of the 1609 measures being considered in this construct only 16 were applicable to Liaison and only 53 to CIS. It was also true that only 59 measures were applicable to IO, but IO has been specifically identified in Joint Pub 3.0 and this area is gaining significant attention, so the question was how to maintain the impact of these measures while eliminating the specific OE area. The solution was to combine CIS with Current Operations (COPS). All COPS areas have a Communications team, and the CIS area directly supports COPS, so this move was logical. The Liaison function is typically carried out by the Commander or the Senior staff, and any Exchange Officers are typically assigned to the Command Element (especially Senior Officers of Flag or General Officer rank) so the Liaison measures were combined with the CMDE OE. In both cases, the skill assignments were valid, so the combination of the CIS and LIA OE into COPS and CMDE respectively was very easily accomplished.

The other area that appeared to dilute the effects of the measures was the skill assignment portion of the methodology. A review of the numbers of selections for each skill assignment was conducted to ensure that the skill assignments selected were being represented in the calculations and having an impact when the observed values were significant. Of the 28 originally designated areas from Table 6 there were eight that did not receive assignments in significant numbers. Three of the skill assignment areas did not appear more than 10 times in the model. Additionally, there were five more that appeared less than 30 times in the model. The one value with 32 occurrences was Timeliness, but this was directly cited in verbiage of the JMET descriptors and was retained. The eight skill assignment areas that were not being utilized did have measures that could have an impact, so a strategy similar to the one adopted for combining CIS and LIA into other areas was adopted for the skill assignment areas. The combination strategy is depicted in Table 39:

Original Assignment	Revised Assignment
Acceptability	Suitability
Balance	Objectivity
Clear Communications	Leadership
Courage	Decisiveness
Flexibility	Adaptability
Liaison	Initiative
Negotiation	Perceptiveness
Trust	Responsibility

Table 39: Skill Area Reassignments

Both instances of combination or reassignment seemed logical, but an independent review of these choices was also required. The very rapid development of models for Major Combat Operations, Noble Resolve, and the Operational Level certification exercise provided the opportunity to validate the choices made in both instances above as well to review the weighting strategies for either individual measures or LDF.

6 ANALYSIS OF RESULTS

6.1 Analysis of Results

The analysis of results provided in subsequent sections of this dissertation focus on key aspects of the research that required resolution to provide results for all three use cases that were useful to the decision makers from the organizations involved. The main emphasis items were performance assessment for FA PANAMAX 07, performance assessment for Noble Resolve 07-2 and TOP OFFICIAL 04, and the costing and ROI efforts for these as well as for the Major Combat Operations use case for U.S. Fleet Forces Command.

6.2 Analysis of Normalized Weighting Approach

6.2.1 Strengths of Normalized Weighting

The most attractive feature (in terms of calculation and effort) of the normalized weighting approach is the ease of calculation. The very simple approach of merely weighting each of the values under consideration equally based upon the total number of values in question simplifies calculations and is easily accomplished. The EXCEL™ spreadsheet tool set can easily adjust to numbers of values using the COUNIT function, and the weight of each observed value can be adjusted with a simple paste link statement. This feature requires initial setup, but even with modifications as experienced in both FA PANAMAX 07 and Noble Resolve 07-2, which consisted almost exclusively of deletions to measures, the normalized weights and resultant averages were calculated with each deletion. The addition of measures, as was required for the FA PANAMAX 07 staff exercise, is just as easily accommodated within EXCEL™.

The normalized weighting method was utilized for every type of calculation throughout this research. That is to say, the normalized averages for skill assignments were averaged within each OE area, and all OE areas were averaged within each APE

function. This approach, especially with respect to the OE groupings is entirely consistent with the method proposed by McGinnis and Stone (8).

6.2.2 Weaknesses of Normalized Weighting

Could this method be used to assess Operational Level performance? The simple and quick answer is yes, but with some reservations. The relatively large numbers of measures being assessed for an event (for this project those numbers ranged from 366 to 692 measures) can be averaged without a significant loss of confidence in the results in most cases. The basis for this assertion, using APE values, is depicted in the Table 36. There is only one case where the weighted average differs from the normalized average by more than 0.03 (the goal for this research, but even that case is within 0.04, which is still useful) which would permit use of either method. The reason for the disagreement in the one instance, The Execute value for FA PANAMAX 07 is that there are only 6 or 7 values being averaged and the value of each has a significant impact. The number is dependent upon whether or not all OE groups are represented within an APE category, and in the case of FA PANAMAX, the FP area has no values, so there are only six factors to consider. An examination of the supporting values for each of the OE groups represented in Table 40 shows the reason for the difference in values between the normalized Averaging and the Floating Weight Averages.

		Total Measures	Norm Avg.	Weighted Avg.
MCO	A	126	0.7063	0.7301
	P	231	0.8026	0.8228
	E	335	0.8044	0.7948
NR 07-2	A	74	0.8434	0.8474
	P	270	0.8393	0.8393
	E	247	0.8378	0.8469
FAPNX	A	65	0.8614	0.8673
	P	130	0.8293	0.8328
	E	176	0.8592	0.8239

Table 40: Averaged Performance Assessment Results

The two major contributors to the overall score would normally be considered to be the Command Element and Current Operations by virtue of the large proportion of the total number of measures in the Execute function. These lower values of assessed performance (either at the T2/T3 break or just above) would lead one to believe that the overall value for execution should be closer to these values than the mid 80's. The normalized averaging approach permits the remaining four OE groups, which comprise only 26%, as shown in Table 41,of the total measures to raise the overall score substantially. In the areas where the values are similar, there are no major differences in the OE grouping values, and thus the results do not differ significantly as seen in Table 40 above.

	Total Measures	Norm Avg.	Floating Wt. Avg.
CMDE	54	0.8258	0.8156
COPS	76	0.8073	0.7975
FOPS	10	0.9100	0.9183
INTEL	13	0.8877	0.8942
IO	16	0.7969	0.8192
LOG	7	0.9214	0.9293

Table 41: OE Group Scores for Execute Function for FA PNX 07

There is one other instance where the results from the Normalized Averaging approach yield a significantly different result than the Floating Weight Average, and that is when there are a relatively low number of measures (for the models discussed this number was typically less than 7 measures) and there were multiple T-4 scores. The Future Operations OE skill assignment area of leadership from the MCO model is an excellent example of the difference between the two methods. The Observed Values for each measure, the weighting factor, and the results of the Normalized Averaging and Floating Weight Averaging methods are shown in Table 42.

Floating Wt.	Normalized Wt.	Observed Value	Floating Wt. Product	Normalized Wt Product
0.2	0.2500	0.85	0.1700	0.2125
0.5	0.2500	0.00	0.0000	0.0000
0.5	0.2500	0.75	0.3750	0.1875
0.2	0.2500	0.85	0.1700	0.2125
			0.5107	0.6125

Table 42: Normalized Values vs. Floating Weight Values

The difference between these values is large, slightly greater than 0.1, which is much greater than the differences in the 0.00 to 0.06 range present in the remainder of the OE grouping. It should be noted that since there were 17 skill assignment groups for the FOPS area, and the LDF for this skill assignment grouping was a 0.200, the overall averages for the FOPS OE were 0.7697 (Normalized Average) and 0.7634 (Floating Weight Average) respectively. Thus, the significant difference for this skill assignment area had no noticeable impact. The fact that the area had a Low LDF is the reason for the lack of impact for the Floating Weight Averaging method result.

6.2.3 Analysis of Approach for the Three Use Cases

The observed behavior throughout all three applications was that large groupings of measures in a skill assignment area, even with as high as almost 1/3 of the measures with scores of 0.75 or less, yielded results that were nearly identical (on the order of 0.003 - 0.007 different) for both the Normalized Averaging and Floating Weight Averaging methods.

The key to deciding whether to use the Normalized Weighting approach as the performance assessment method is whether it is permissible to consider all tasks, all skill assignment areas, or all OE groupings as equally important. Although this requires the least amount of effort by any staff, it is doubtful that such a proposal would be

acceptable. A possibility would be if there were limited numbers of metrics, such as in the isolated case of the RDD event measures devised for the Territory of Guam. There were only 47 measures used in both Noble Resolve (NR) 07-2 and TOPOFF 04 (originally 49 measures were devised, but two were declared not applicable by the observers in Guam). The proficiency of the Staff and the Emergency Responders was impressive in both exercises, and with relatively few observed values below a T-3 rating in either exercise (7 at T-3 in NR 07-2, 1 at T-4 in TOPPOFF) the results for Normalized Averaging were never more than 0.024 in magnitude. Even in that particular case, the two values were equally spaced on either side of the T-2/T-3 rating break point, thus stating that this area (Execute in the APE assessment) was at T-2 would not be inappropriate. The weightings of various measures were not uniform, so there could be concern if a large number of very low observed values had been entered. A test was conducted in the Execute APE function where there were only 11 total measures. Seven of the 11 were assigned values in the T-3 or T-4 range. All T-4 values were 0.00 (4 total) to ensure that the impact would be noted. In this case the values of the Normalized Averages and Floating Weight Averages still agreed within 0.024, and with the overall score in the 0.45 range, it is not likely that the lack of agreement between the averaging methods would be as noteworthy as the overall performance which could only be judged as substandard.

A review of the Normalized Averaging method for the MCO, Noble Resolve, FA PANAMAX, and RDD specific scenarios does show that for the 83 calculations of performance (whether for a specific OE, of APE as a result of multiplication by AHP factors, or for APE from OE groupings), 81 (97.6%) of the results differed by less than 0.03. This would be acceptable for any staff evaluation.

The values that exceeded 0.03 are displayed in Table 43below:

Model	OE/APE	Difference
FAPNX 07	FP	0.0708
MCO	LOG	0.0562

Table 43: Performance Assessment Differences greater than 0.03

The logical question is whether or not there are any commonalities that could account for these specific instances of discrepancy. In both instances above, the discrepancy is a direct result of the difference between Normalized Averaging and Floating Weight Averaging.

The Future Plans discrepancy from FA PANAMAX 07 is because there are very few measures, only 13 in fact. Of these measures, two were assessed as T-4, one value being a 0.00, and one was assessed as T-3. The 0.00 assessment was part of the Skill Assignment group Thoroughness, and the weight for this particular measure was 0.8 (High). The resulting Floating Weight value for this area was 0.1700, far less than the 0.4250 value from the Normalized Method. When the 0.1700 value was combined with the other area values, the discrepancy noted in Table 39 occurred. If the assessed value for this measure was even as high as 0.65, a valid T-4 value, the difference for these calculations would have been less than 0.03. The impact of a HIGH importance measure, especially when there are very few total measures, is readily apparent in this example.

A similar situation was observed for the MCO case. In the LOG OE area, the Preparedness Skill Assignment area had two 0.00 values, and a T-3 value. In this case, both 0.00 values were associated with measures of HIGH (.8) importance. The T-3 value was associated with a measure of LOW (0.2) importance. The resulting totals for the Normalized Average and Floating Weight Average were quite different, 0.6700 and 0.6161 respectively. The additional factor relating to the overall difference was that this Skill Assignment area was the only HIGH (0.8) value area for the OE, so this very low value caused a significant impact on the overall assessment for LOGISITICS, but that would be the reason for the weighting factors. There were 48 total measures for this area

split between 11 Skill Assignment areas. Only one other area received an assessment of less than 0.800, so the averaging method hid the most important area's discrepancy.

The preceding discussion would tend to support the notion that the Floating Weight averaging approach would be a preferable approach to an normalized averaging method, but the results also show that the Normalized approach is consistent and reliable. The researcher did not expect to find that the Normalized method would be as effective as shown, and does believe that this approach could be used in an instance where simplicity or time constraints were a factor.

6.3 **Analysis of Floating Weighting Approach**

6.3.1 Development of Approach

The discussion in the previous section has already provided some insight into the use of the Floating Weight Averaging approach. The Logarithmic Driving Function method devised by Dr. Will Startin (23) relies on a weighted averaging method to arrive at the values for each Skill Assignment area. As these sums were being calculated for each area or each OE group within the APE method, the agreement between the two methods was useful in checking the validity of the normalized averaging approach as well as formulating a factor for the LDF method.

The possibility of using these values as an alternate method to calculate the overall performance of any specific OE or for the APE method came from the realization that the LDF factors (High/Medium/Low) could be multiplied by each of the weighted averaging factors from the LDF equation and dividing the sum of those values by the total of all LDF factors. The formulation is an extension of equation 2.4 (23):

$$S_i(T_j) = \frac{\sum w_j T_j}{\sum w_k} \tag{6.1}$$

where S_i = i^{th} pillar, applicable to the execution of a specific skill assignment area or OE grouping

T_j = j^{th} metric in the i^{th} JMET-process bin, representing observed measure data

w_j = factor between 0 and 1 representing the relevance of the metric T_j to pillar S_i

It is relevant to point out that this approach is very similar, albeit utilizing weighted averages, to the approach proposed by McGinnis and Stone (8). The aggregation of the Skill Assignment areas or OE groupings is carried out exactly as is done within each Skill Assignment area or OE grouping, and thus there is no requirement that the sum of all the weighting factors equals unity. Using this method to calculate an overall performance parameter for each OE or APE function did serve the purpose of providing a useful check of the Normalized Averaging values, especially early in the development of the performance assessment methodology. It was not until the development of the FA PANAMAX, Noble Resolve, and MCO use that the potential utility of this approach was realized.

6.3.2 Strengths of Floating Weighting Approach

The major advantage of the Floating Weight approach and the compliment to the LDF method is that it permits (in fact requires) the training audience to assign an importance to each subtask. The training audience must also decide either by defaulting to the number of JMETs which represent a specific Skill Assignment area within any given OE group or OE grouping within the APE functional group, or by conscious choice, which of these areas is most critical to staff performance. The operational level staff involved with the development of the performance criteria for FA PANAMAX 07 appeared to be very comfortable assigning weighting factors indicating the importance of each subtask. They also appeared to be comfortable, even with the wide variance in the staff's assessment team experience, with the methodology and actual assignment of the

LDF weights for both the Skill Assignment areas and OE groupings within the APE functional areas.

The LDF assignment method in Section 4.3.4 was based upon the numbers of measures in each Skill Assignment area or OE grouping. This approach appeared reasonable absent any specific direction by observed Training Audiences.

The three opportunities presented by Noble Resolve 07-2, FA PANAMAX 07, and TOPOFF 04 allowed the researcher to validate the JMET driven approach to assigning LDF weighting factors, and as a consequence further refine and validate the Floating Weight Averaging method.

For both Noble Resolve 07-2 and TOPOFF 04, the team executing either the experiment (JFCOM J-9 for Noble Resolve) or the exercise directors on the Territory of Guam were given multiple opportunities to review the JMETs selected, the weighting factors applied to the measures, the skill assignment areas, and finally the LDF weighting factors.

The effort to reduce the JMETs, which did not apply to the Homeland Defense scenario, as well as the addition of the Radioactive Dispersal Device JMETs, were significant alterations to the original performance assessment template and were concurred with by J-9 personnel. The use of the Skill Assignment area totals and OE grouping totals as the guide to assigning LDF weighting factors was seen as the most reasonable method to emphasize the importance of those areas. The weighting factors for each measure were reviewed and agreed to by the same J-9 Directorate personnel.

The RDD event in TOPOFF 04 provided the opportunity for both J-9 and Guam exercise personnel to review the RDD JMET performance assessment template with the understanding that the LVC exercise would also utilize the same JMET measures. As shown in sections 4.4.9 and 5.4.1, both staffs agreed to reduce the number of measures by two, in both cases because the scenario did not support those measures but did not alter the weights, the LDF weights, or T-rating standards. As a result, the performance

observations for the RDD event in both Noble Resolve 07-2 and TOPOFF 04 were applied in the same fashion. An overview of the key results for the two events shows some differences which are candidates for further investigation. Those key results are provided in Table 44:

	Norm. Avg.	Fl. Wt. Avg.	Norm. Avg.	Fl. Wt. Avg.
	NR 07-2	NR 07-2	TOPOFF 04	TOPOFF 04
CMDE	0.8645	0.8642	0.8963	0.9053
COPS	0.8186	0.8308	0.9014	0.9080
LOG	0.8599	0.8553	0.9217	0.9238
Assess (AHP)	0.8401	0.8440	0.9050	0.9113
Plan (AHP)	0.8452	0.8472	0.9016	0.9088
Execute (AHP)	0.8369	0.8424	0.9170	0.9204
Assess	0.8434	0.8474	0.9500	0.9500
Plan	0.8393	0.8393	0.9389	0.9330
Execute	0.8378	0.8469	0.9017	0.8986

Table 44: Key RDD Performance Assessment Results

The two noticeable results that would suggest possible further analysis are both differences between the aggregated OE results using the AHP factors and the directly calculated APE values. The first instance is in the Plan area from TOPOFF 04 (viewing the Normalized Averaging results) and the second is in the Assess area from TOPOFF 04.

In the case of the Plan Normalized Weighted average value of 0.9016 as compared to the value of 0.9389 for the AHP linked OE results; the reason for the difference appears to be the fact that of the 11 measures, nine were assessed as T-1 during the direct APE calculation. The CMDE OE grouping only has two measures, so the 0.9500 result of the calculation only deserves a LDF weight of 0.2 which does not significantly impact the overall assessment value. There are seven T-1 observations in

the LOG OE, and these do have an impact on the calculated result, making it 0.9217. The value shown in Table 40 of 0.9389 appears to be a good representation of the Staff's ability to plan for RDD events. The AHP factors for this case do not permit the very strong performance in Planning in the Logistics OE to have the impact that just considering planning measures does in the direct calculation. It is also worth noting that in either instance, the values are in the T-1 range, the difference is less than 0.04, and the difference would probably not be seen as significant by a Commander.

There are only 47 total measures for the RDD performance assessment tool, and 18 of these are split between Assess (7) and Execute (11). As in the discussion in section 6.1.3 where a difference between the Normalized Averaging and Floating Weight Averaging was evident where the weighting factors and low numbers of measures had a significant impact, there would appear to be some similar relationship in these instances.

The difference between the OE linked value for APE and the directly calculated value is straight forward in the TOPOFF 04 instance. There are only 7 Assess measures, and these were all assessed as T1. No matter what method is used, the value for Assess will be 0.95. The contribution to each OE Skill Assignment area is not as significant. For the CMDE Skill Assignment area of Competency, there are three Assess measures, all weighted at 0.5. Even though this Skill Assignment area is the most important in the CMDE OE, the overall value for Competency is 0.9352, but the final assessed score for CMDE is 0.9053. The relatively low importance of Assess as a part of the CMDE area contributes less than a 0.95 value to the linked total.

Although both the COPS and LOG OE groups have two Assess measures, these measures are part of the Situational Awareness Skill Area, and there are even numbers of High and Medium weights. For COPS, there is one T-4 measure, therefore even though the Situational Awareness area has only one measure less than T-1 (the resultant value for the Skill Assignment area is 0.9159), the final assessed score for the COPS area is 0.9080. Furthermore, the Situational Awareness Skill Assignment area within LOG has only 0.95 observed values, but this area has a LDF weight of 0.2, therefore even though the final assessed score of the LOG area is 0.9238, there is no AHP relationship which

can transform all three OE areas with values of less than 0.95 into a value of 0.95. In this instance, the very high observed values for Assessment cause a difference in values between the OE and APE cases, but the difference is still less than 0.04 which is close to the goal for agreement for the model discussed in Chapter 3.

The Floating Weight Averaging method does portray performance in a manner that accommodates the importance of the measure weights, the importance of the Skill Assignment areas, and the OE groupings. The results are not greatly different than those from the Normalized Averaging approach, but when there are significant differences, the differences do reflect the areas which need either improvement or where great proficiency has been attained.

6.3.3 Weaknesses of Floating Weighting Approach

The basis for the advantage of the Floating Weight Averaging approach is also the greatest detractor for use. Until there is a well established methodology or template for assigning weighting factors to performance measures, as well as the same processes for assigning LDF weighting factors, the Floating Weight Averaging approach suffers from the fact that it requires a significant amount of time and involvement from the staff being assessed. Initially these issues can be overcome by either accepting the time commitment or by utilizing personnel to develop the performance assessment tool with Operational Level experience.

A second disadvantage of this approach is the fact that every scenario will require some modification of weighting factors, so there is a necessary amount of staff involvement and time investment for any changes of scenario required for certification or proficiency. The changes are necessary if the Commander wishes to use this approach, because even if measures apply in two different scenarios, the training and proficiency goals of the operational level staff may be vastly different. An excellent example of this situation is the weighting factors which apply to establishing liaisons with Host Nations, Country teams, other US Government Agencies, and Non-Governmental Organizations. The weight for this measure (OP 4.7.2 Conduct Civil Military Operations in the Joint Operations Area, Measure 3) is 0.8 (High) in a scenario such as FA PANAMAX 07,

since the Mission Essential tasks support operations which support primarily Stability Operations and FHA/DR. If the scenario was oriented towards Major Combat Operations, measures of this ilk would be assigned 0.2 (Low) factors until the latter stages of combat operations or in the post-combat phase. There are many similar examples, but the Operational Level staff must make these decisions if the Floating Weight approach is to be valid.

One variation of the Floating Weight Averaging method was to use only the values for either Skill Assignment areas that had corresponding High or Medium LDF weighting factors when calculating the overall value for a particular OE area or APE functional area. Although this approach appeared to have potential, the great variation in results (in three of seven instances greater than 0.03 different from using all values), shown in Table 45 below from the FA PANAMAX data as an example, did not make this approach viable as an alternative to using all the values within each OE group or APE area:

	Fl. Wt. Avg.	Limited Fl. Wt. Avg.
CMDE	0.8314	0.8277
COPS	0.8205	0.8514
FOPS	0.8888	0.9401
FP	0.6836	0.5483
INTEL	0.8732	0.9428
IO	0.8359	0.8188
LOG	0.9060	0.8912

Table 45: Comparison of Fl. Wt. Avg. and Limited Fl. Wt. Avg. for OE areas

It is worth noting that in the three areas where the values exceeded 0.03, two (FOPS and INTEL) were due to very strong performances and the High factors dominated the

calculations, and in the case of FP, the performance had been weak, and the High factors reflect that as well. As a macro approach, this would be acceptable, and with some modification, this approach could be useful.

6.3.4 Analysis of Approach for the Three Use Cases

The first indication of the potential significance of Floating Weight averaging was realized when examining the results of the FP assessment during FA PANAMAX 07. This discrepancy was also noted in Section 6.1.3. Further analysis of the results of all the models did provide a basis for the differences in this approach from the values obtained using Normalized Averaging. During the brief to the Operational Level Commander after the conclusion of FA PANAMAX 07, the results of the FP area were discussed in detail and the Commander agreed that the Floating Weight approach was more useful in pinpointing areas requiring additional emphasis.

The final observation arising from analyzing this approach was the fact that regardless of the model constructed, the number of measures which apply to Assessment is generally between 12.5% and 18.2%. The best example of how many measures apply to Assessment is from the FA PANAMAX 07 performance assessment tool since the measures were devised by the Operational Level staff. Even in this case, only 65 of 371 (17.5%) applied to the Assessment area. The impact of these low numbers was not an issue for either averaging method, but the fact that Operational Level staffs rely on Assessment as the key to correcting Plans and adapting to changing situations, this appears to imply that more Assess type measures should be developed. The increasing emphasis on effects may assist in developing more Assess type measures.

6.4 **Analysis of Logarithmic Driving Factor (LDF) Approach**

6.4.1 Difficulties Encountered with Unrestrained LDFs

A major goal of this research effort was to attempt utilization of the Logarithmic Driving Function approach for performance assessment. As discussed in Chapter 3, the principle for this approach appeared to not only have utility for accurately calculating the observed performance, but could also provide very powerful insight into either superior

performance which may be a best practice, above standards or poor performance which requires immediate remediation.

Several methods of assigning values for the LDF factors were attempted during the early research effort, primarily while building the initial instantiation of the performance assessment tool. The very marked difference between the values noted for calculating performance for the OE areas and the APE models was of concern. Though the LDF method appeared to have validity for the APE performance calculations, it was typically inaccurate for most OE performance calculations, unless there were a very limited number of factors. A representative set of results for this early work is portrayed in Table 46. None of the results achieved would have been acceptable for any operational assessment.

	LDF	Limited LDF	Fl. Wt. Avg.
A	0.5998	0.7212	0.8286
P	0.5660	0.7144	0.8283
E	0.5629	0.7168	0.8244

Table 46: Sample Values for LDF using a Linear Formulation Approach

6.4.2 Use of High/Low/Medium Values for LDF

The desire to utilize the LDF approach for performance assessment drove the researcher to attempt one other LDF weighting assignment approach. This approach was not recognized until formulating the results for FA PANAMAX and was inspired by a conversation with Dr. Startin (23). Dr. Startin had used values of 0.8/0.5/0.2 as his H/M/L values, and these did result in better approximations for performance. As one extension, to further minimize the impact of areas of Low importance, the researcher used the values 0.65/0.35/0.075 respectively for the H/M/L values for LDF weighting factors. Early in the research, such an assignment had been made, but the impact of approach was not recognized because this attempt was made prior to viewing any use cases where performance assessment was reviewed by outside parities. Table 47 below

shows the impact of the two H/M/L approaches on the APE functional areas as contrasted with the Floating Weight averages. Both of these approaches were marked improvements over an of the earlier assignment approaches.

	LDF [36]	LDF[37]	LDF[38]	Fl. Wt. Avg.
A	0.7059	0.7362	0.8390	0.8284
P	0.6722	0.7060	0.7918	0.8245
E	0.6937	0.7220	0.8471	0.8239

Table 47: Comparison of LDF Values with Varying H/M/L values [36] [37] [38]

Although the values obtained for each APE area using these revised LDF weights were still not within the 0.03 range obtained for the Normalized Averaging and Floating Weight Averaging calculations, these results were, as previously noted, far better than any obtained in earlier attempts. The potential to use only the High or Medium values for LDF (Limited LDF) factors also appeared to have potential, although the fact that the result might be higher than the averaged value now had to be considered. The values observed for Limited LDF for the FA PANAMAX, Noble Resolve, RDD, and MCO use cases are all displayed in Appendix K.

The LDF approach for FA PANAMAX was very encouraging and based upon previous results from previous use cases, somewhat surprising. In this case, regardless of the LDF value calculated for any APE function, the magnitudes of the final values were relatively the same as the Floating Weight averages. In fact, for seven of twelve, the

[36] The H/M/L values used here are the .8/.5/.2 suggested by Dr. Startin

[37] The H/M/L values used here are .65/.35/.075 as determined by the Researcher

[38] The H/M/L values used here were as a result of AHP calculation

difference between the LDF value and the Floating Weight average was less than the 0.03 goal. Three instances were greater than 0.06 different, and two were still at 0.09. These results were encouraging since the LDF method was as accurate as the either averaging method nearly 60% of the time, which had not been typical in the other use cases.

The results for the FA PANAMAX use case led the researcher to revisit the LDF approach for the MCO and Noble Resolve use cases, including just the RDD assessment. The values for the LDF approach presented results that were completely unexpected. The MCO model, the Noble Resolve model, and the RDD model for Noble Resolve were uniform. The Assess LDF value, using either H/M/L format, was consistently 0.15 to .27 lower than the Floating Weight average for the Assess value. The Plan and Execute values for the MCO, Noble Resolve 07, and the RDD model for Noble Resolve were not nearly as disparate. In only one instance (the Execute LDF value for Noble Resolve day 2) was a value as much as 0.10 different from the Floating Weight average. The values generally differed between 0.03 and 0.05, close to the desired 0.03 difference desired.

The use of the limited LDF approach was also investigated as a possible alternative to the use of all the LDF factors. The results, as noted in Appendix K were better than those from the LDF method and did not display the same puzzling issues with the Assess performance values as seen in the previously discussed cases. The use of this method does appear to have merit, but the ability to consistently choose the correct Skill Assignment areas or appropriate OE groupings does appear to have a significant impact on the results. The researcher used only those Skill Assignment areas (within each OE) or the OE groupings (within the APE functions) which had a High or Medium LDF weighting factor (regardless of which H/M/L approach was used). An examination of the values in Appendix K reveals that in many instances the Limited LDF value exceeds (previous cases were always lower in value) the value for the Floating Weight average. It is noteworthy that the Execute and Assess areas of the APE functional area most often had the Limited LDF value exceeding the Floating Weight Averages. Even though there were values that exceeded the Floating Weight Average values, the validity of this approach was surprising. Of the 42 comparisons made from the Appendix K, only 2

Limited LDF values exceeded a difference of 0.06, and were not greater than 0.09 from the Floating Weight average. Twenty six values (61.9%) were within the 0.03 of the Floating Weight averages, which was the desired goal. There were 14 values greater than 0.03, but less than 0.06 different from the Floating Weight average values. It would appear that even if the puzzling results of the LDF method could not be easily resolved, the Limited LDF approach would be suitable.

One additional method of linking the performance observed within the OE groupings to APE was suggested by Dr. Startin (69), which was the second level aggregation approach. The reasoning was that if the Skill Assignment areas within each OE grouping could be aggregated, this might be a more appropriate method of linking the contributions of each OE groupings within each APE calculation than simply using the AHP values with the results of the OE values. The results observed for the FA PANAMAX use case were somewhat puzzling.

The Assess LDF results often significantly deviated from the Floating Weight average values and were not of the same relative magnitude as the Plan or Execute results (in the APE calculations). When the second level aggregation was implemented, the behavior reversed in almost every case where the Assess value had been low from the previous discussion.

The apparent cause for the change in values can be explained due to the numbers of measures observed in the Assess area. The relatively low numbers of Assess measures for all models translated into fewer factors being multiplied to produce the terms for the second level aggregation. There were never more than 18% of the measures for any model which applied to the Assess area. For all but the RDD model, the Plan and Execute measures were typically double or triple the number of Assess factors, so the total number of terms involved would certainly cause the products for either to be smaller than the Assess product.

One excursion to attempt to resolve the Plan and Execute second level aggregation was conducted with the FA PANAMAX 07 model. In this experiment, only

terms which had LDF weights of 0.5 or 0.8 were used. The result was improved, both the Assess and Execute values were of the same magnitude (the difference with the Floating Weight average was less than 0.10 in both cases) as the Floating Weight average result. Even though the Plan results improved, it was still not useful.

It is therefore apparent that additional research into the possibility of having the second level aggregation more accurately portray the contributions of each OE group to the APE functional areas is merited. It should be possible to have both LDF methods as options for performance assessment reporting.

6.4.3 Use of AHP Factors

The issues with the LDF method results discussed in the previous section, as well as the overall results of the Operational Level staff exercise, were briefed to the Operational Level Commander (75). One of the people present at the briefing was Dr. Startin, who expressed an interest in the results. Dr. Startin examined the results from the Operational Level exercise and a meeting was arranged between the researcher and Dr. Startin (74). When examining the results, Dr. Startin noted that the use of the AHP method appeared to be successful, and he asked if there had been any thought to directly applying the AHP values as the LDF weighting factors. He noted that since the Staff had chosen the measures, and had concurred with both the Skill Assignment area designations and APE designations, that this could be seen as having surveyed the staff. The suggestion did seem reasonable, and several of the models (with the exception of the Initial Bins model) were modified so that the APE Floating Weight averages and LDF values would be the result of AHP derived LDF factors. Where second level aggregations had been conducted, the AHP factors would also be applied as the LDF weights.

In every instance the AHP values observed for LDF, Limited LDF, and second level aggregation increased. The increase in value appeared to be promising, especially since the previous use of the H/M/L results had been improving with a shift from the 0.8/0.5/0.2 application to the 0.65/0.35/0.075 application. A closer examination of the results was not as encouraging however. Six performance assessment cases were tested

for the LDF and Limited LDF results, while four cases were tested for the second level aggregation results.

The relatively good agreement noted in the previous sections when comparing the LDF, Limited LDF, and second level aggregation to the Floating Weight averages (with a goal of a 0.03 difference) was not repeated with the AHP values as LDF weighting factors. Although 13 of 18 results for LDF were within 0.06 of the Floating Weight average APE values, only four of these values were within 0.03. The Limited LDF results were slightly better, with 15 of 18 values being less than 0.06 of the Floating Weight average values, but only two values were within 0.03. The second level aggregation results were quite poor, with eight of twelve values exceeding the Floating Weight average values by 0.1.

6.4.4 Advantages of this Approach for Performance Assessment

The LDF approach does reflect poor performance when an area of High importance (whether within a OE group or within an APE group) occurs. This is also true if there are poor performances for multiple Medium importance areas -- whereas poor performance in an area of Low importance is not significant. These were all desired characteristics of the LDF approach.

The major issue with the LDF approach is the fact that large numbers of factors are utilized. As discussed in the previous sections, whether in the Skill Assignment areas or when attempting to calculate the results for a second level aggregation, large numbers of factors tend to result in a very low value. The APE calculations never involve more than seven factors and these values for LDF and Limited LDF were consistently better, though the limited LDF factors consisted of only the High or Medium LDF weight OE groupings.

The utility and potential of the LDF approach, as well as the possibility of arriving at an APE value by applying a second level aggregation, would still seem to be the ultimate performance assessment calculation methodology. The results obtained throughout this research are promising, but with the exception of the Limited LDF

approach, the LDF method is still not consistent enough to be used as the basis for reporting results or as the primary basis for assessing the performance of a staff.

6.5 Use of Analytical Hierarchy Processing to Link Performance Assessments

6.5.1 AHP and the Realities of Operational Level Staffs

The issues of deciding what tasks were important and how each of the Organizational Elements contributed to the overall missions of an Operational Headquarters was something that required resolution. A simple survey method was something that could apply, but how could the results of such a survey be translated to resolve the issues with Operational Level staffs? The best answer appeared to be AHP. Ernest Forman delineates the three basic principles of AHP:

- Decomposition – Structuring a complex problem in to a hierarchy of clusters, sub-clusters, sub-sub clusters and so on.

- Comparative Judgments – The construction of pair wise comparisons of all combinations of elements in a cluster with respect to the parent of the cluster.

- Hierarchic Composition (Synthesis) – Multiplication of the local priorities of elements in a cluster by the 'global' priority of the parent element, producing global priorities throughout the hierarchy and then adding the global priorities for the lowest level elements. (40)

These principles were completely aligned with the approach the researcher required as well as matching the structure of the JMETs and the hierarchical nature of an Operational Level (or any) staff. The AHP method appeared to be the optimal method for resolving the issues of prioritization of missions for the Operational Level staff as well as providing a potential means of linking Organizational Element performance to the Operational Level missions of Assess, Plan, and Execute.

Further research into the AHP approach increased the confidence that this method would apply since the tasks performed by an Operational Level staff do have the characteristics Forman described as necessary.

Some elementary attempts to use the method were successful, but the issue that quickly became apparent was, how could all the members of the Operational Level staff be effectively surveyed to gather the data necessary to compare OE functions to each other in each APE context? There are over 100 members of the staff, and the experience levels of the members of the staff vary significantly. Large variations in experience and knowledge could make the results, even if fully collected, inaccurate. Further, there would be a need at some point to "average" the inputs to obtain a single view of each Organizational Element's contribution to the Operational Level staff mission to permit a pair wise comparison. The optimal way to "average" the opinions of more than 100 staff members was somewhat daunting.

One method to solve this problem would be to only survey the most senior members of the staff, which would eliminate the experience and perspective issues, but surveying the staff would require time on the part of the Staff and the researcher. Even this approach would require some sort of "averaging" although the opinions would probably not have the very wide variations as would be observed on the entire staff. This approach may be the most desirable, but it was not practical for this particular staff nor would it be for any large operational level staff.

The solution seemed to depend on some way of determining what Skill Assignment areas would be most important within any OE grouping as well as a method of determining which OE groupings would be most important within any APE functional area. All the model constructs, regardless of number of measures, could be decomposed into Organizational Elements, APE functional areas, and further decomposed into Skill Assignment areas. The Operational Level staff's decision to use Training Objectives that had been collectively devised presented an interesting opportunity.

Since the JMETs that comprised the certification requirements were well defined, the Operational Level staff was able to determine what tasks and sub-tasks were required observation as Training Objectives. The tasks and sub-tasks were submitted by each Organizational Element in a format which was similar to the descriptive language utilized within the UJTL. The Staff had debated which tasks must be completed so that the final

list was, in fact, a defacto survey of the importance of each Organizational Element's contribution to the overall mission.

The researcher approached the Operational Level staff Training Coordinator with the following proposal: Would the Staff agree to review the weighting of each sub-task requiring completion, would the Staff review the Skill Area assignments, and would the Staff review the LDF weighting (which would be the importance of Skill Assignment areas within the OE construct as well as the OE importance within the APE construct)? The proposal was accepted, and the "Rosetta Stone" described in Section 4.2.4 was created and reviewed by the various Organizational Element personnel responsible for performance assessment.

The results of the staff review were described in detail in Sections 4.2.4 and 4.2.5 as well as in Section 5.4.1.2. Once the sub-task weights were assigned, the Skill Assignments were assigned, and the APE functional areas were assigned, the ability to assign importance values to either the Skill Assignment areas or APE appeared to be an issue related to the number of measures. The researcher used ratios of measure numbers within each APE category as the method of assigning LDF weights therein. The potential to do the same with each Skill Assignment area within an OE group also exists, but time constraints for the overall staff performance precluded constructing the matrices necessary to use this approach effectively.

6.5.2 Modification based upon Applied Measures

Using the process for all the required assignments necessary to assure proper functioning of the performance assessment tool, the researcher tested the AHP methods on some earlier models. The two logical candidates for testing were the initial instantiation and the MCO use case. The initial instantiation model was also modified so that the LDF weights were assigned by the H/M/L protocol.[39]

---

[39] Both the .8/.5/.2 and .65/.35/.075 values were used in the models to assure consistency. The key results are available in Appendix K.

The performance linkage between the OE and APE calculations was discussed in Sections 5.3. The validity of the calculations was re-enforced by the results observed for the performance assessment during Noble Resolve 07-2 (Section 5.4). The researcher decided to use the JMET driven[40] approach for linking OE results to APE results exclusively for the FA PANAMAX 07 performance assessment.

6.5.3 Results of AHP Linkage

There were two issues to be resolved regarding use of AHP. Both had been directly addressed by Forman or had not been encountered or considered likely in the development of the method. One of the caveats for using AHP was that the clusters or any sub-clusters needed to be able to be related to each other within one quantum level. Thus, in pair wise comparison, one cluster could be up to 9 times as important as another cluster, but the goal was to have the elements relate so that there were not wide disparities in the clusters. The initial instantiation and MCO cases did fill this requirement when looking specifically at the A, P, E differences since all were numbers between 126 and 709. The breakouts for the individual OE groupings within the APE functional areas did not meet this requirement, since sometimes there were over 100 measures for one OE grouping and as few as 4 for another. Both the Noble Resolve and FA PANAMAX APE measures had the roughly the same percentage of Assess measures as the Initial Bins and MCO models, but as previously noted, the total numbers were in the 65 – 74 range, and thus the measures were not of the same order of magnitude as the Plan and Execute measures. The ratios were not greatly different however, that is to say, the numbers of Plan to Assess measures or the numbers of Execute to Assess measures were not greatly different in any of the models as depicted in Table 48 below:

[40] JMET driven refers to using the numbers of JMETs relative to each other within Skill Assignment areas or APE functional areas as the determining factor for importance.

	Initial Instantiation	MCO	Noble Resolve	FA PANAMAX	RDD
Plan to Assess Ratio	1.83	2.55	3.65	2.0	1.5
Execute to Assess Ratio	2.66	2.23	3.33	2.7	4.14

Table 48: Ratios of APE measures for Applicable Models

There was still the issue of the applicability of AHP when there were vastly different numbers of measures within a Skill Assignment area or within an APE functional area, but as previously discussed, performing the AHP method on the sub-clusters of the Skill Assignment areas was not pursued due to time constraints. In general, the numbers of measures within any OE met the guidelines of Forman, so returning to the Skill Assignment areas to apply AHP should pose little difficulty.

Even with the sometimes occurring disparities in numbers, the AHP method was applied to the Initial Bins and MCO models for linking the OE performance assessment results to each APE functional area. As discussed throughout Chapter Five, the results were consistent and the results were very often within the +/- 0.03 difference goal for the research. Since the FA PANAMAX model had far fewer measures than either of the cited examples, the AHP method was implemented.

The remaining issue was related to circumstances where there were NO measures that had to be dealt with. This was not addressed by Forman, but attempting to multiply anything by 0 would not be useful, and since 0 would also necessarily be a divisor in the matrix, the method would not function.

The researcher's resolution was to assign a value of unity to any area (which only applied to Future Plans measures during Execution) that had no measures. The lack of measures for an OE grouping had not been previously encountered since the JMETs had

been taken from the pre-existing list. There had been instances of very low numbers of JMETs within an OE grouping, but no complete absence of measures. The ultimate output of the AHP method is a set of factors which add up to unity. Using a value of unity, or even having widely disparate numbers of measures did not cause the totals to be less than unity. Additional investigation revealed that reducing the value of the absent measures to a value such as 0.1 or 0.01 had no significant impact on the factors. (The variation was not seen until the third or fourth decimal place).

The ability to link OE performance to APE performance for comparison was successful as discussed in the previous chapter. Also discussed was the potential to use the AHP values as the LDF weight factors. The results for this use of the AHP results were not quite as successful, but the improvement in the APE values from previous H/M/L was promising.

A final area requiring additional research was discovered when the researcher was using the AHP method to relate the three OE grouping performance measures to the APE functional areas for the RDD model for both Noble Resolve and TOPOFF 04. There were no Planning measures that applied to COPS for the RDD scenario. There were also only 11 total measures. Using a value of unity for COPS did result in a sum for all factors of 1, but the values for the LDF weights varied greatly when values of either 0.1 or 0.01 were substituted for unity in the matrix for the number of COPS measures. As noted in the discussion for FA PANAMAX, the disagreement between potential LDF weights was still in the third decimal place, but since the effort to use 0.01 as the COPS value was no more difficult than using 0.1, the former was used. Calculations using both LDF weights were identical to the fifth decimal place. Calculations using the either H/M/L set of values were also identical with the exception of the Floating Weight Average values, which varied in the third decimal place. (The AHP values were 0.93264, the 0.8 H/M/L approach value was 0.93303, and the 0.65 H/M/L value was 0.93098. All would be acceptable as results).

6.6 Analysis of Joint Mission Essential Task Assignment to Costs

6.6.1 Ability to Assign Costs to JMETs

When this research was initially conceived, the goal and optimal situation was to demonstrate the ability to assign costs to specific JMETs since these could be related to events within the Operational Level in a "scale neutral" fashion.[41] The difference in training scenarios, the potential differences between purely synthetic and LVC events, and the variation in exercise types posed some possible constraints on achieving that goal.

A potentially more granular method of assigning costs would be to assign costs against a specific measure within any given JMET. Though this would be the most specific way to assign costs, that level of specificity was determined to be impractical very early in the costing development phase of the research. Costs ended up being aggregated under specific JMETs so that the total resources expended were of a significant enough magnitude to be useful.

The other two possible means of assigning costs were either against a specific function, which would be in the APE grouping or against a skill assignment area. Though there was some validity for this approach, as the various models were constructed, the ability to compare costs between models using either of these approaches also proved impractical.

Discussions with U.S. Fleet Forces Command (USFFC) indicated that using the JMETs would be useful since the request for personnel and resources were related to the JMET basis of any specific exercise. Further, accreditation of USFFC training sites is tied to JMETs, so knowing what could be accomplished in terms of JMETs was the preferable way of viewing costs.

---

[41] The term scale neutral is used to describe the situation where the size or number of participating units, staffs, or personnel is not important. The only consideration is the accomplishment of the JMETs at a performance level which is considered Operationally Ready.

The initial effort to link costs to JMETs was discussed in Sections 4.5, 4.6 as well as in Chapter 5. The original cost assignments were made to the most specific level of JMET possible and did not prove to be difficult. An example of this type of assignment would be the costs incurred due to the participation of U.S. Army Patriot batteries. These units function exclusively as Theater Air Defense units, so the JMET which applies is OP 6.1.5 (Conduct Joint Operations Area Missile Defense) (1). There were also functions carried out by U.S. Air Force units and U.S. Navy units which were similar, but as previously discussed, there were other areas of Air Defense that applied, so relating events as was done for FST events might not be useful if other air defense tasks under the more general OP 6.1 Provide Operational Air, Space, and Missile Defense JMET were conducted and costs were assigned.

6.6.2 Reliance on the Parent JMETs

The immaturity of the cost assignment process required that baseline protocols for costing be implemented. The more generic "two-digit"[42] JMETs were easily understood by the personnel at USFFC, and did address the Training directorate's need for discussing resource requirements with JFCOM. Based on this understanding, the MCO model costs were assigned at the two-digit level. The ability to discern where resources were not expended between events, even with completely common scenarios, was not hampered by using two-digit JMETs. The example was discussed in Section 5.2.2. Since costs had to be assigned for the LVC event to permit ship movement, but none were required for the FST event, the ROI calculation was based solely on similar JMETs. This point was discussed with the Operational Level Commander during the FA PANAMAX 07 debrief (74). The Commander was particularly interested in the consistency of the JMETs used to compare the MCO LVC vs. synthetic application and was satisfied with the approach adopted by the researcher.

---

[42] JMETs are set up in a hierarchical fashion. OP 6 sets forth the Operational Tasks which address Provide Operational Force Protection. OP 6.1 and OP 6.1.5 are as described. It is common to address JMETs as two-digit or three-digit, or as parents (two-digit) or children (three-digit).

Until the cost assignment process is well understood and staffs prepare for either exercises or experiments specifically targeting JMETs as the activities being considered, the use of two-digit JMETs will be sufficient for comparisons and as a method to compare exercises between CJTFs.

6.6.3 Use of JMETS beyond Two Digits

The ability to achieve the granularity and thus confidence that the resources allocated are directly impacting readiness is not a distant or intangible goal. Some of the common missions which are already readily accepted as mission areas in any Joint context include:

- OP 1.1.3 Conduct Joint Reception, Staging, Onward Movement, and Integration (RSOI) in the Joint Operations Area
- OP 2.2.3 Collect and assess Meteorological and Oceanographic (METOC) Operational Information
- OP 3.1.9 Conduct Dynamic Targeting
- OP 4.4.3 Provide for Health Services in the Joint Operations Area
- OP 4.7.3 Provide Support to DOD and Other Government Agencies
- OP 4.7.5 Coordinate Politico-Military Support
- OP 4.7.8 Establish Disaster Control Measures
- OP 5.1.2 Manage Means of Communicating Operational Information
- OP 5.5.1 Develop a Joint force Command and Control Structure

- OP 6.1.5 Conduct Joint Operations Area Missile Defense
- OP 6.2.9 Coordinate Personnel Recovery
- OP 6.2.14 Employ Operations Security (OPSEC) in the Joint Operations Area
- OP 6.3.4 Protect Information Systems in the Joint Operations Area (1)

The listing above is by no means exhaustive, but it is noteworthy that for both the MCO model and FA PANAMAX 07 exercise all of these JMETs were part of the supporting JMET structure either required by the COCOM or desired for Operational competency. The Noble Resolve scenario contained 11 of these JMETs; the two that were not included were those concerning targeting and personnel recovery.

A very important point regarding all of these JMETs is that they all require specialized equipment, specific training elements, specifically trained observers, and specialized scenario support. They are ideal candidates for near term implementation of granular activity costing. This listing is general enough yet specifies necessary capabilities for any Operational Level Commander carrying out the responsibilities of a CJTF since none of the listed JMETs are necessarily restricted to Major Combat Operations. The MHQ w/MOC Concept of Operations document (68) addresses all of the cited JMETs as capabilities that are required for any Numbered Fleet Commander (the Navy Operational Level Commander) to be accredited as either a Joint Force Maritime Component Commander, a Joint Force Naval Component Commander, or in some cases a CJTF. A review of the requested training for Unified Endeavor 06-1 also shows that the five of the cited JMETs were specifically requested.

Since the first Operational Level certification is complete and included all the cited JMETs, specifically targeting these for costing for future events would not only provide a greater ability to view specific costs, it would permit establishing a more flexible baseline for all costs within an evolutionary activities based cost methodology.

6.6.4 Operational Level Certification Exercise Model Analysis

The cost data supplied by the Operational Level staff met the requirements stipulated by the Commander in that all costs were captured in a manner such that both JMET assignment and cost categories could be viewed as desired. Though there are no preceding events, projecting the costs which can be avoided in follow on events and the costs which must be accounted for in all future events is easily accomplished. The indirect costs and one-time/event specific costs are excellent candidates for avoidance and thus a projection for ROI is possible with the current data.

Now that a certification event has been completed, it should also be possible to examine the very large costs necessary for event support from JFCOM and make a determination whether that same level of resources will be necessary for future events. If some of those costs can be avoided, another form of ROI is available since the JMET requirements for certification are identified.

The Operational Level staff admitted that the very fast paced preparation for the event did not permit gaining a better understanding of the exact nature of the ROI project. Armed with the ability to view specific JMET requirements for any future events, cost data can be properly categorized and appropriate JMETs assigned as event planning is being undertaken.

The cost categorization was well done by the Staff. The researcher's initial impression of the data was that JMET assignment would be more difficult than for the MCO or RDD case, but since the listing of JMETs was very specific, this impression was very quickly proven incorrect. As with the MCO case, some of the costs were very readily assigned due to the very specialized nature of the costs, such as Logistics and Foreign Disclosure Officer costs. The staff felt that some of the costs should be handled in an even distribution since the certification JMETs were addressed throughout the scenario.

This type of exercise would be an excellent candidate to utilize three-digit JMETs in all future events. The scenario is very specific, and events such as Joint Personnel

Recovery require specific hardware, training support, and network capabilities are well understood. If these specifics had been identified for FA PANAMAX 07, assigning costs for ROI purposes would have been easily achieved.

6.6.5 MCO Use Case Analysis

The MCO model, as noted above could easily be extended to include several three-digit JMETs. At this writing, the limited set used for cost comparison and to establish the capability to directly link costs to accomplished training was more than sufficient. The major objectives of the FST series exercises will continue to coincide with the objectives of the Joint Task Force Exercise, especially at the CFMCC level. Changes to the CFMCC requirements can easily be integrated into both exercises so increasing either the number of JMETs addressed, or expanding the JMETs already being executed to permit better cost visibility and fidelity will not invalidate the approach.

The listing provided for the MCO model already can be used to compare events of the same type (e.g. FST to FST) either year-to-year or event-to-event (as was shown in Section 5.3.2). The current methodology also permits the comparison of LVC vs. synthetic events since both can be viewed in a scale-neutral sense. The JMETs chosen can permit costs to be tracked in conjunction with the JMETs that JFCOM desires for use during Joint National Training Capability (JNTC) events, which would also permit comparison to other similar events on a JMET by JMET basis.

The only issue encountered with the USFFC data was the fact that no indirect or one-time/event specific data had been captured. All other data was transmitted in the categories requested and was easily converted into the most appropriate JMETs.

6.6.6 Noble Resolve Use Case Analysis

Noble Resolve's RDD event provided an excellent basis for JMET assignment to costs. There 10 JMETs comprising the "core" JMETs for this experiment were established as the result of the desired training objectives of the participating agencies, the overarching goals for the experimentation program, and a series of Warfighter Challenge issues which relate the J-9 experimentation program to JFCOM. The very

clear direction for establishing the JMET baseline greatly enhanced the ability to capture the costs required to develop and execute either a LVC or synthetic scenario.

The apportionment of costs for Noble Resolve (NR) 07-2 was discussed in Section 4.6.5, which reduced the ambiguity which might have existed in assigning costs to the JMETs associated with the RDD. Most of the apportionment issues for costs were very straight forward and easily validated by J-9 personnel. Any comparisons using the costing data for NR 07-2 which may be conducted for future events with a similar JMET basis (with 10 JMETs, it should be possible to compare a significant portion if not all JMETs for future LVC or synthetic events) will only provide an initial estimate of ROI if a similar or more formalized cost assignment methodology is used (since the JMET cost alignment for NR 07-2 was not specifically discussed during experiment formulation or execution).

If the costs required to execute the RDD event in TOPOFF 04 can be captured at some future date, then a comparison similar to that completed with the MCO model would be easily executed. At this writing, the LVC costs are not available, so a thorough comparison is not possible.

Conversion of the standards in the Homeland Security Universal Task List to allow easy comparison to the JMETs in the UJTL would permit capturing cost data for the Noble Resolve series in a manner that could be beneficial to the Federal, State, and Municipal levels, which would also demonstrate ROI to each of these agencies. It would also provide a more formal vehicle for collecting cost data that would more completely characterize the costs required to execute this series of experiments. The conversion would also permit exercise planners within the J-7 directorate visibility into key mission essential tasks that are common to both military and non-military organizations, and permit ROI calculations for both types of organizations.

6.7 **Analysis of ROI Results**

6.7.1 Operational Level Certification ROI Analysis

The detailed cost information provided by the Operational Level staff permitted the most thorough examination of ROI potential relative to the other opportunities available. The potential for cost avoidance, as well as the ability to use this event as a template for similar Operational Level staffs (especially those of the same Service), is not yet fully understood. Certainly the ability to use the JMET data as a basis for follow on events is obvious, but if costs were to be retrospectively assigned to three-digit JMETs so that a more detailed comparison was possible for future events, then a situation similar to that forecast for USFFC would be possible. The very well categorized data and the very specific JMET listing for certification are the enablers for useful future ROI discussions.

Details which could be modified include more specific parsing of the costs for the support personnel in terms of specific JMET accomplishment. This would not impact the methodology, it would only change the percentages of costs assigned to specific JMETs. Additional opportunities to capture JMET based activities costing are extensive.

The myriad of Net Present Value (NPV) and Return on Invested Capital (ROIC) discussions germane to the methodology developed herein could generate significant interest during throughout the Programming, Planning and Budgeting process

6.7.2 MCO ROI Analysis

One of the objectives of this research effort was to be able to portray ROI based upon performance and costs in an environment neutral manner. The use of the UJTL as the common basis for costs and assessment also provides a scale neutral basis for ROI as well.

There are some costs which are not accounted for in the MCO model, notably the costs associated with the aviation which directly support the Operational Level Commander's requirements. This is important since the costs to utilize the range facilities, set up or move targets, and to assure environmental standards are being

carefully documented. Since the costs for aviation are not specifically detailed by three-digit JMETs in a synthetic environment, the ROI calculations which were carried out are viable. These show the tremendous advantage for synthetic training, especially at the Operational Level vis-à-vis a LVC environment.

The comparison of FST events is also valid since a similar JMET baseline was used to assign completed training and costs incurred. As discussed in the previous section, the ability to detail costs to the three-digit level would not be difficult and this could be done retrospectively as long as both the LVC and synthetic event JMETs were assigned consistently.

The ROI calculations permit discussion of two significant issues: 1) Which aviation costs merit inclusion in the LVC scenarios and 2) What can be done to more effectively or efficiently deliver synthetic training. Both of these issues were raised while debriefing FA PANAMAX to the Operational Level Commander, which was the basis for his direction for very detailed cost figures and JMET assignment.

US FFC does have this data and has the methodology used to arrive at the results so that future event decisions can be made in a manner that could permit achieving greater granularity and possibly avoiding some costs. Knowledge of the types of costs, especially in the indirect and one-time/special event categories, as events are being planned and executed should permit more formal cost collection data and more specific JMET assignments.

6.7.3 Noble Resolve 07-2 ROI Analysis

The very specific nature of the personnel cost parsing, based upon the factors discussed in Section 6.6.6 have enabled the JFCOM J-9 to review the proportion of resources allocated against the core JMETs. The potential to conduct an ROI estimate between the RDD events in both NR 07-2 and TOPOFF 04 (which would only comprise three JMETs) is a very obvious first target since most of these costs should be available from the Territory of Guam. There may well be costs that are associated with some individual measures within the cited OP 7 JMETs, but unless these costs comprise some

significant portion of the costs expended on Guam (either of the same order of magnitude or a significant fraction such as greater than 10%), neglecting these costs would still permit calculation of an initial LVC vs. synthetic ROI comparison.

The costs for the entire TOPOFF 04 exercise, since the scenario was the same as NR 07-2, would also provide an excellent opportunity to calculate a LVC vs. synthetic ROI. Care would have to be taken that those elements which were not a part of NR 07-2, such as the RDD event in Phoenix, AZ, must be excluded for the ROI to be meaningful, but this again could be the basis of an initial estimate of the relative value of the LVC event to the synthetic event. Much as with the employment of tactical units for the U.S. Military, perhaps the ROI for large mobilizations of First Responders is not as significant since there might be long periods of inactivity as is sometimes experienced by tactical military units. In these cases, perhaps limited use of these types of capabilities could be designed to permit some training, but only if there is utility for both the response units and the higher level government offices. The analogy between the Operational Level and tactical level and inefficiency of attempting detailed training for both simultaneously is evident and should receive the same consideration.

6.8 **Sensitivity Analysis**

6.8.1 Sensitivity Analysis Framework

The desire to quantify a mixture of soft and hard measures to provide a numerical value for performance can introduce errors that would cast doubt on the accuracy of the performance assessment. Several factors could introduce error or possible variation in the performance assessment results. A framework to assure any operational level staff of the validity of the results was not developed during this research effort, but several of the factors most likely to introduce error, specifically the weighting factors, binary results, and skill assignments were examined to verify their impacts.

One area that was not considered in terms of sensitivity analysis was the actual observed scores. The staffs in question wanted to use a four tiered scale, and also wanted the ability to assign unsatisfactory grades with values ranging from 0 to 0.69 if the

situation warranted. The only instance where observed values were inflexible was in the assignment of the binary (Yes/NO) evaluations. In the instance of a binary score, a Yes received a value of 0.85, or a mid-T2 rating, since satisfactory completion of the task would be expected of any staff. An unsatisfactory answer was assigned a value of 0.50, easily in the range of 0 to 0.69, but not a 0.0 value. The reasoning for this approach was discussed in Section 2.1.2 and was accepted by all the staffs involved in this research. Any unsatisfactory score did result in a noticeable decrease in any skill assignment area, and for those Organizational Elements such as Future Plans with very low numbers of measures, unsatisfactory observations had a significant impact on overall performance as would be expected.

The final number of skill assignment areas, as discussed in Section 5.5, was reduced from the initial 29 to 20. Though this number of skill assignments adequately represented the desires of the staffs involved, some of the skill assignment areas such as decisiveness, efficiency, and projection may be combined with other skill assignment areas in the fashion discussed in Section 5.5 so that the measures can be retained, giving greater impact to the measures than they currently have in some situations.

A simple method to determine the change in impact, carried out as a test case, was done with the Command Element results from FA PANAMAX. If the three skill assignment areas are reassigned there is some variation in the results, but it is not significant as depicted in Table 49 below.

	N. Avg	Lim N. Avg	Fl. Wt. Avg	Lim Fl. Wt. Avg
Original Skill Assignments	0.8062	0.8304	0.8291	0.8277
Reduced Skill Assignments	0.8367	0.8304	0.8314	0.8277

Table 49 Sample Variation of OE Results

The results for the Floating Weight Average, the preferred method of calculating performance for all the staffs, did not vary enough to warrant concern, but that may not be the case in some instances where there are fewer measures, so further examination of the need or desire to retain the 20 skill assignment areas for not only military but other applications is warranted and sensitivity analyses should be conducted for those skill assignment areas that appear to have little influence.

The final set of values which merit examination in the operational level staff context are the weighting factors. A discussion of the variation of weighting factors and an example of the A, P, E values using three different weighting factors was displayed in Section 6.6.2 and Table 47. It certainly appears that there is merit to investigating the impacts of the weighting scheme used for the LDF factors; since the LDF values more closely approximated the Floating Weight Average values as different schemes (changing H/M/L schemes or using AHP derived factors) were implemented.

Is the same true for the variations of H/M/L weighting factors for individual subtasks? There must be a way to differentiate the importance of subtasks, as discussed in Section 4.3.3, and all the staffs involved in the research felt that a simple H/M/L scheme was the least confusing and most easily agreed to by all members of the staff. Thus, the effects of variation need to be examined within the constraints of each of those regions.

Three different cases of variation were examined as examples of a framework for any assessment model. The assessment areas of interest from FA PANAMAX were Future Plans and Logistics since these areas had observed values on the border of T3 or T1 respectively. In addition any of the A,P,E areas could be of interest, and the Execute area was selected since the results for this area were lower than had been expected by the staff.

A simple program was devised in EXCEL™ which permitted varying each of the factors between the boundaries used in the models, for example varying the High value

randomly in the range from 0.65 to 0.80. The observed values were input as had been recorded by the staff. The results in Table 50 show that in all but one instance, the case of the lower H/M/L values in the FP case, the variation of weights did not change the results significantly.

.8/.5/.2	0.6836	0.8912	0.8166
.65/.35/.075	0.6417	0.9060	0.8234
Variation	0.7020	0.8960	0.8306

Table 50: Effects of Varying Weights

The very low numbers of measures for FP, and several low grades in areas with higher importance were the cause for the variation, but the range of scores observed validates that the overall impact of weighting appears to be correct. The variation method used ran 200 iterations and the standard deviation for any of the areas was never more than 0.005, which would not impact any overall score for an operational level staff.

One further variation was implemented to observe the possible changes if the ranges were expanded to an even greater degree. In this variation the ranges were expanded to nearly 0.3, so that a high weighting factor could vary from 0.6 to 0.9. The only change noted was that the FP observed value was 0.8984, which is not significant; the other area values were unchanged.

The very limited example cited here does provide insight that should be incorporated into any discussion with a staff. The impact of weighting factors and the selection of those factors, the number of measures for each area, and the variations that might occur should be fully explained and understood, thus reducing unexpected results at the conclusion of the exercise.

There is always discussion as to whether or not the right values for T1 through T4 have been assigned. Though it would be more difficult, a similar trial could be conducted to portray the impact of such changes to any staff prior to commencing an exercise.

7 CONCLUSIONS/RECOMMENDATIONS FOR FURTHER RESEARCH

7.1 Conclusions

7.1.1 Not the Panacea, but a Methodology to Build Upon

The goal of this research was to develop and demonstrate the feasibility of a methodology that could be reliable and repeatable for assessing operational performance, linking the costs required to achieve that performance by use of a common language, and finally to provide several potential formulations of Return –on-Investment (ROI) which could be used to aid decision makers at the Operational Level.

The research evolved such that the performance results and more importantly the costing results were "scale neutral." The scale neutrality is critical because many of the discussions of ROI have traditionally centered on the numbers of personnel trained, or the fact that some percentage of Joint Mission Essential Tasks (JMETs) were addressed by some or all of the personnel trained. The effectiveness of that training was captured only in estimates (and in some cases still is) of the number of successfully accomplished JMETs compared to the total number that were involved in the particular scenario involved. This is a "scale neutral" measure as well, but still doesn't address these key questions:

- How ready is the Operational Level Staff to perform the JMETs which apply to the eventual theater of operations?
- What capability does the Operational Level Staff have to Plan, Execute, and Assess the Missions assigned?
- What will it cost to accomplish the above and what is being achieved for that cost?

The research also attempted, successfully in the view of the researcher, to maintain an "environmentally neutral" focus on performance assessment. If the methodology developed cannot be used in an operational environment to maintain or ascertain proficiency, in a LVC environment, and in a completely synthetic environment, then it is of very little use to any Operational Level Commander. Further, the JMETs and supporting tasks must be those performed in the operational environment so that any additional resources requested, remediation required, or training mandated can be stated in the same terms as the JMETs necessary to successfully perform in the operational theater.

The example of ROI for the Major Combat Operations model excluded JMETs which did not apply to both a LVC and synthetic environment. Additionally, though the application of aviation costs would be more insightful regarding the maximum ROI attainable between LVC and synthetic events, since no significant additional costs are incurred when employing constructive adversaries, and the results of engagements, strikes, and any politico-military responses are conducted synthetically in either environment, as long as the aviation costs are consistently excluded, the ROI calculations will remain valid.

The methodologies that evolved throughout the course of this research for both performance assessment and costing linkage have been utilized by personnel on Operational Level Staffs, within the J-9 Directorate of JFCOM, and by exercise directors in Guam with a minimum of training successfully. Additionally, the coincident products, the "Rosetta Stone" and Observer Grade Sheets have been successfully employed by the same personnel.

One possible concern that must be addressed is the validity of the results, especially from the point of view of those who either reviewed the results, or desired to use the results for official purposes. There are examples of validity in each of the use cases, which provide both an internal, the view of the staff in using the process, and external, either as an official means of reporting or in a request for additional work, which can be cited.

The personnel at JFCOM (J-9) as discussed in Section 5.4.1 furnished the metrics devised in the course of this research to personnel in the Territory of Guam for use in determining the effectiveness of Radiological Dispersal Device actions. During TOPOFF 04, JFCOM (J-9) personnel observing the reports from the Territory of Guam noted that the measures reported, and the standards established, permitted the Territory to easily meet the requirements of the exercise, they were the only organization (of three involved in a RDD scenario) to meet all time and reporting standards. As was noted in Section 5.4.1, no changes were made to the measures, so the validity of the measures for training within the Territory of Guam and for use in a national setting was verified.

The training directorate personnel at U.S. Fleet Forces Command (USFFC) reviewed the data as noted in Section 5.3. The areas of particular interest to the staff were the potential application of three-digit JMETs to achieve more granular cost data, and the ability to view costs in terms of JMETs. The potential for this methodology as it could apply to other Numbered Fleet Staffs as well as to assess the capability of each of the Navy's nine Maritime Headquarters (MHQ) that have a Maritime Operations Center (MOC) mission is compelling enough that USFFC has requested additional work in both areas. There is a significant correlation between the required tasks and costs associated with either a limited Joint Task Force Commander (CJTF) certification or the required tasks and costs associated with accrediting a MHQ w/MOC. The ability to use the methodology and report readiness on a more frequent basis, thus permitting a more frequent ability to assess each staff's readiness is a desired outcome of the task. Additionally, there is the potential to show ROI in terms of reduced training time, reduced resource allocation, and greater reuse of scenarios for similar mission sets thus reducing staff time required for exercise design or preparation.

Commander Second Fleet (C2F) personnel have had the most significant opportunity to validate the methodology. In addition to using the results of FA PANAMAX 07 to report the staff's readiness to deploy as a certified CJTF, the costing and ROI data by JMET category was reported to SOUTHERN COMMAND, the Combatant Commander (COCOM) for C2F. The performance assessment methodology

has been used for a second exercise, BLUE ADVANCE 08, to assess the proficiency of the staff after a five month period. The exercise is similar to FA PANAMAX, but having a greater understanding of the types of missions that are required in a HA/DR and Stability Operations environment permitted the staff to scrutinize the Training Objectives and make some changes in standards, weighting factors, and applicable JMETs. These have been assessed, but the results have not been reported to the commander.

The Commander also requested that the cost data be captured so that the ability to show ROI, especially since almost all Indirect and One-Time/Event Specific costs could be eliminated, would be available for reporting to the COCOM. The ability to impact the resource allocations so that areas that were considered deficient during FA PANAMAX could be more appropriately funded was also a goal. The effectiveness of targeting resources is also not known. Finally, the staff realized that too much emphasis had been placed on Execution, and there was a shift of emphasis to Assessment, such that a bit more than 22% of the measures to be assessed during BLUE ADVANCE would be in the Assess category, an increase of 5%. This shift caused Plan and Assess to be slightly more than 50% of all the measures assessed, not as high as the staff would like, but a trend now possible due to the repeated use of the methodology.

There are issues to be resolved for both methods, and those will be discussed further later in this Chapter, but the initial methodology appears to provide a feasible approach for computing ROI and does meet the goals of reliability and repeatability. Unfortunately, like many other activities of this sort, until the use of these methodologies is consistent, the real utility will not be realized.

The reason that a very proprietary software was not utilized was that many of the methods currently available to the military are either extremely difficult to use, requiring large numbers of support contractors, or tend to operate with a series of software patches because the original intent of the application did not address military operations. The intent was to utilize a tool such as EXCEL ™ that is readily understood, especially by the middle level members of most staffs (who tend to be very adept at such software applications) so that as familiarity with the mathematical formulas and procedures was

attained, improvements could be readily made and implemented, enhancing the performance of the tools developed.

There is no reason that either methodology could not be operated in a classified medium, but again, to gain maximum exposure and acceptance, the project was specifically maintained at the UNCLASSIFIED level. As long as the JMETs used are not tied to specifically Classified Operational Plans, the tools can be used in an unclassified medium and the data collected for costs as well as the ROI calculations completed can be briefed, discussed, and then acted upon in any Operational Level or decision making forum.

The methodology, especially the approach to weighting factors, skill assignment areas, and cost parsing are merely points of departure. They are based upon operational experience and the initial values and formulations of the three previously mentioned elements have been validated by five separate teams of subject matter experts, but the final state of all of the key elements and the optimal employment of the tools will require repeated use. Expectations that either solutions or clear choices will be immediately available are unrealistic and must be avoided.

7.1.2 Major Obstacles to Implementation

Both aspects of this research effort, performance assessment and the linking of costs to performance, have obstacles that are primarily cultural that must be overcome or at least resolved for the methodologies to be useful. The researcher has encountered both during the course of the research. It has also been the experience of the researcher that with some explanation and negotiation, the methodologies have been accepted, even if in a limited form, which has proved their utility.

The first cultural obstacle is the perception that assessing performance is either that such an effort somehow usurps the authority and ability of an Operational Level Commander to determine the Staff's Operational readiness. An additional aspect of this view is that those who perform the assessment are somehow violating a trust or

attempting some form of purely mathematical construct reminiscent of the early 1960's with Secretary of Defense McNamara.

It is the duty of the Operational Level Commander (or any Commander) to realistically assess the readiness and operational performance of the staff or unit he/she commands. Although there are numerous methods available to assess tactical level training from an individual up to a complex organization such as a Carrier Strike Group, Infantry Division, or Wing, there is no common method for assessing Operational Level performance. This methodology was developed specifically to begin to fill that void.

The methodology, as implemented by the Operational Level staff during a major training exercise, gave the staff complete control of the tasks and supporting tasks selected to accomplish their specified JMETs. The staff had the ability to choose the Organizational Elements (OE) that would accomplish the tasks while choosing the weighting of the tasks with the scenario in mind. The use of Skill Assignment areas permitted the staff to view how the tasks would aggregate for ultimately calculating the performance of both the OE groups and the Assess, Plan, Execute APE) functional areas while also being able to choose the importance of each Skill Assignment area's importance or OE groups importance relative to any APE functional area. The latter set of choices were based upon the numbers of measures chosen and the ratio of each Skill Assignment area's number of measures compared to the total for any specific OE or any Organizational Elements total number of measures compared to the total for any APE functional area. The use of Analytical Hierarchy Processing to determine the proper relationship of each OE to the each of the APE functional areas proved successful so that a direct comparison of each method of assessment could be accomplished.

The use of either Normalized Weighting or Floating Weight Averaging as methods to calculate performance was successful and the decision as to which method to utilize can be made by the Operational Level staff. The use of the Logarithmic Driving Function (LDF) method for assessing performance was partially successful, and when used with only a limited number of factors, does provide results that compare favorably with either averaging method. The LDF method does very quickly portray the impact of

a significant deficiency in performance, especially when used with only limited factors since the result is ultimately a product of factors directly related to any specific Skill Assignment area or OE grouping.

The freedom provided to the Operational Level staff to make all of the choices detailed above did not preclude assigning JMETs to each measure or from reporting into the Defense Readiness Reporting System (DRRS). In fact, the results from the exercise discussed were used without modification. DRRS requires different formatting to input results, and the results must be grouped in JMET order, but the use of the translation tool developed in the course of this research permitted those manipulations and calculations to be completed in less than an hour. The issue of a more effective reporting method will be discussed later.

Two other successes that were achieved in the course of this research were the ability to develop standards that are useable in very short periods of time. The majority of the JMETs in the UTJL do not have published standards that are applied evenly throughout the services or Combatant Commanders. There are approximately 2700 potential measures cited in the UJTL, only about one-third of these have had standards developed and published. About half of the JMETs chosen by the Operational Level staff and all of the 49 JMETs required for the Radioactive Dispersal Device event for Noble Resolve (NR) 07-2 required development. Only 15 measures (from a total of 461) for both of these events were changed by assessors. The researcher acknowledges that the background he possesses provided great insight, but once the attribute of the measure is decided upon, then the specific ratings to assess performance to show the appropriate level of proficiency was not difficult to ascertain.

The final area demonstrated as a success was the ability to determine which warfighting skills were most applicable to the measures. The selection of the characteristics that can describe the qualities necessary for a Staff and Commander to make critical decisions and direct forces was enabled by the language of the UJTL. The basic set of 20 characteristics such as decisiveness, competency, situational awareness, adaptability, preparedness, and thoroughness are either used with the UJTL descriptions

or the measures suggest those attributes. Using these characteristics or Skill Assignment areas as the groupings for measures, the ability to indirectly measure these essential qualities of a dynamic organization and leadership team becomes possible. As long as the characteristics are consistently applied, the results do mirror the performance of a staff. These Skill Assignment areas apply not only to military organizations, but any large organization, especially those working in an operational environment such as federal, state, or municipal emergency operations staffs or large corporate management headquarters with global operations.

Although the resistance to performance assessment was expected, effective implementation of the methodology, especially when the staff involved was given extensive choice in the process, was achievable and successful. The more surprising and in retrospect, the more difficult challenge was to overcome the inability to quantify or perhaps gain visibility on costing regarding training and readiness.

There are multiple persons with bits of cost data; there are just as many methods in place to track costs. Each instance where costs were requested received a different reaction. There was complete openness and the desire to explicitly delineate costs to the greatest detail possible. There was a desire to examine costs, but with only a general framework, and there was the inability to get cost data. The ability to obtain, categorize, and then assign JMET correspondence to cost data evolved throughout the research.

The most difficult of these steps is the categorization of costs. Comprehending the concept of either direct or personnel costs was not an issue, but the ability to comprehend what was meant by indirect costs, or the need to capture one-time/event specific costs proved more challenging. Indirect costs are not well understood, and there is a misconception that since military personnel don't "produce" things the way industry might, that there is no loss of productivity. In fact, many exercises at the operational level require changes in communications capabilities, changes in daily functionality, or the need to add communications nodes to maintain the normal channels of communications while the inherent capabilities are being used for exercise purposes. The resources for these indirect costs are rarely anticipated or accounted for. Similarly, the

unexpected costs which seem to arise all too frequently due to the need for licensing software or specific hardware suites to permit an exercise to continue have not been captured regularly or with any fidelity. These types of costs contribute a far more significant percentage of the overall costs for readiness training than is currently understood. If these costs can be captured, assigned JMETs, and then viewed for ROI purposes, significant cost avoidance should be possible.

Once costs are categorized, if there is a "core" set of two-digit (and eventually three-digit) JMETs which costs should be assigned against, the assignments are not difficult. Explicit training objectives, overarching guidance, and training audience priorities drive the generation of training scenarios, the flow of the scenario, and the time required to achieve the desired objectives, thus assigning costs in proportion to these guidelines is very straight forward. This was the case with both FA PANAMAX 07 and Noble Resolve 07-2. Both had many of the elements discussed above. The assignment of costs, either in a cooperative fashion or for submission was completed in less than two hours in both cases. The key was that both staffs had categorized the costs in advance and had very specific JMET baselines to which costs could be assigned.

The ability to move to even more granular three-digit JMETs was introduced in Section 6.5.2. These JMETs are not the definitive listing or a constraint, merely the most common three-digit JMETs observed in the course of this research and those which can easily have costs assigned. Using a combination of two and three-digit JMETs in the cost assignment process will enable all planners, but more importantly Commanders and decision makers with visibility on what training has been effectively accomplished and for what cost. This can be the true inception of quantifying the "bang for the buck" or the "So what?" question that must be answered consistently and accurately in order to justify the very high costs associated with Operational readiness and training.

7.1.3 Needed: A Reliable and Methodical Approach for Decision Making

The ability to portray costs, assess performance and calculate the ROI associated with Operational readiness and training have been goals of not only the Department of Defense but any large Federal, State, Municipal, and even major corporate organizations

for many years. The intangible nature of some of the activities carried out in these primarily service oriented organizations have always made the assessment of performance elusive. Associating costs to training is readily understood in terms of equipment operation in an industrial setting or even at the tactical level of warfare, but at the Operational Level, though the UJTL provides a framework, implementing that framework in a meaningful way has been inconsistent at best.

The pressure to justify the expenditure of resources on personnel, equipment, and readiness will not abate. There are situations wherein the need to allocate resources in large amounts very quickly cannot be avoided or anticipated. If a methodology to show why those costs were necessary and the basis for comparing the situation and requirements exists, then the reason for the decision to expend the resources can be clearly articulated.

7.1.3.1 Within Operational Level and Headquarters Commands

The continually expanding need for CJTF Headquarters to deploy on very short notice to support a wide range of military operations from disaster relief to stability operations does not appear to be decreasing. Commanders need a methodology to prepare the Staff for the most likely missions and tasks expected, a way to assess the readiness, so that assistance can be targeted for maximum effect, and a way to express the time and resources which will be required to confidently state that the CJTF Headquarters is ready to deploy.

Maintaining the performance assessment tool provided in this research and using it to ascertain the state of readiness of the staff can provide a baseline to request training. The JMETs required by each Combatant Commander are well defined, and the ability to match personnel to specific JMETs has been accomplished within the framework of the Operational Level Commander certification and FA PANAMAX exercise. The range of JMETs for which the Operational Level staff has demonstrated proficiency is documented in DRRS, and each member of the staff is assigned duties in the necessary board, center, cell, or bureau with the attendant JMETs assigned. It is merely a matter of entering the observed value of performance during an exercise or to enter an appropriate

value indicating the lack of proficiency for newly reporting members to establish the Staff's current readiness. The deficiencies can now be identified by specific JMET and appropriate training can be either devised or requested to address the deficiencies.

There is also the desire to improve processes by using new hardware, software, or procedural methods. Since the proficiency of the staff can be calculated at any given time, or the level of performance that is required is known for any specific JMET or range of JMETs that any of the proposed solutions are purported to address, determining the veracity of the claim of performance can be evaluated in an objective manner. If the hardware, software, or procedure assures Operationally ready performance for a lower cost, more rapidly, or using fewer personnel, then it can be considered. If not, the additional cost (in addition to any already expended to develop the capability) to achieve the desired performance can be weighed against the well known costs required to achieve the desired performance in the past.

This is not an exhaustive listing, but the potential to accomplish even these rudimentary steps in a quantifiable manner instead of making decisions not founded upon operational performance and a linkage to the required JMETS is now available to Operational Level commanders.

7.1.3.2 Within the Department of Defense (DoD)

The extensibility of this methodology to all of the DoD is probably presumptuous at this point in time. It is not presumptuous to state that if all Operational Level Commands in all the Services adopted the JMET framework for determining each Command's core missions and were to formally approach the costs required to either attain Operational readiness or maintain proficiency, then the common costs for these staffs would be a compelling argument for acquiring systems, developing training scenarios, constructing distance learning courses to address knowledge shortfall, and perhaps even a reason to consider the adoption of capabilities available from serious gaming to rapidly train personnel. All of these examples have several characteristics that can be implemented within the methodology developed during this research. The JMETs required can be readily identified or measures, attributes, and standards that will support

assessing any JMETs necessary to support the mission set can be quickly developed. The costs associated with scenario development, acquiring systems, constructing distance learning courses, or implementing serious games are finite. The ROI for all of these is easily calculated. More importantly, the ability to AVOID costs due to inexperience or training for the wrong mission set can be readily identified.

Again, this is only one potential application of the methodology that could be employed throughout DoD, but it is a very simply achieved first step that can lead to much better allocation of resources and decision making.

7.1.3.3 Within Other Non-DoD Activities

The key to using the methodology began with the search for a common language to assess performance and to link costs to that performance. For Operational Level Commanders in the military, the Universal Joint Task List (UJTL) is the common language. What about other organizations? As shown in the case of the Operational Level Commander desiring to properly define the training objectives necessary to attain certification, even starting without reference to the UJTL is NOT an obstacle. The staff was able to eventually match all the training objectives to the UJTL, but the training objectives were developed to accomplish the Mission Essential Tasks (MET) necessary to attain certification. Military staffs are not the only organizations that have METs. ANY federal, state, or municipal organization that must use teams to provide command and control for emergency situations, whether those are due to natural disasters, terrorist action, or some failure of critical infrastructure require the same skill sets and capabilities as an operational level commander.

The Homeland Security Universal Task List (64) has well defined tasks, and instead of using the APE format, it would be easy to use the Prevent, Protect, Respond, and Recover (P2R2) categories to determine the effectiveness of staff OE actions. The OE groups will have slightly different names and may have functions that are not quite equivalent to those discussed and portrayed in this research, but extending the concepts developed is easily accomplished.

The same issues that must be addressed with the military operational level staff must be addressed in any of these non-military organizations.

- What are the Mission Essential Tasks (MET)?
- Which OE group performs them?
- How does each MET align to the P2R2 function?
- What measures will be used to assess each MET?
- What attribute will apply to the measures selected?
- What are the key Skills that are needed and how are Skills Assigned to each measure?
- How does the Staff wish to assign the importance of the Skill Assignment areas to each OE grouping?
- How does the Staff wish to assign the importance of each OE groupings to each P2R2 functional area?

Those are precisely the same steps used in the development of the performance assessment tool for a military operational level staff, and they apply with equal validity to all the organizations mentioned above as well as many corporations.

The approach developed for cost assignment is equally applicable, and has a similar list of issues which can be addressed:

- What are the core METs which represent the training objectives or operational readiness desired?
- Which category of cost best describes each resource utilized?

- Which METs are most appropriate for the resources allocated?
- What percentage of time, preparation, or effort was required for each MET or were resources expended that did not contribute to any MET?

Once these issues are addressed, and if they are addressed at the beginning of the training process, the ability to view the expenditure of resources against METs is easily calculated. The same principles for ROI, comparison from previous events, or as a method to implement cost avoidance or improve effectiveness in future training events which were discussed in this research apply.

7.2 **Recommendations for Application of Methodology**

7.2.1 Blended Approach Potential

The ability to use either the Normalized Averaging or Floating Weight Averaging methods combined with Analytic Hierarchy Processing (AHP) as the method to link the performance of the OE group performance to APE performance is viable. The issue really becomes a matter of choice for the Commander in question. Since all tasks are NOT of equal importance, and the weight of each measure supporting the tasks can vary with the operational situation, the use of the Floating Weight Averaging method would seem to be the most promising.

For this method to become truly effective, continued work on the development of weights for the measures must be accomplished. The AHP method not only links the OE grouping performance results to the APE results, it also appears to provide effective LDF weighting factors for use in the overall Floating Weight calculations.

When only the Medium or High importance LDF terms are used to determine the LDF performance, the results are typically in alignment with either averaging method, so this approach could be used and would very readily provide the insight as to which areas of performance were causing degradation.

The use of the AHP values for LDF weighting factors can be applied in the same method as the High and Medium factors since the relative weights of each OE grouping do translate into the High or Medium importance categories. Either approach, as well as the use of the variation of the alternative H/M/L values (.65/.35/.075 vice .8/.5/.2) is valid and can be used with the tools provided from this research to assess performance.

The use of AHP as the linking method and as a method to calculate LDF factors may not be the ultimate or even most effective means to determine weighting factors, but it did demonstrate that a method which utilized the inputs of the staff to determine which tasks or organizational elements were most important, without a survey approach could provide valid results. Other methods which utilize a similar thought process may provide even better results, and should be pursued. The AHP method is effective and easily applied when there are relatively low numbers of factors but proves to be somewhat unwieldy when using large numbers as was seen in some of the Organizational Elements. It is also not a practical method for determining the importance of subtasks. Though a simple High/Medium/Low method seems unsophisticated, it is easily understood and used by staffs, so no change to that approach is recommended.

7.2.2 Potential Impacts for Defense Readiness Reporting System (DRRS)

The current implementation of the DRRS does not accept the results produced by this research directly. This shortfall was addressed during the FA PANAMAX 07 debrief with the Operational Level Commander with the recommendation that the developers of the DRRS be contacted to raise the issue. The current requirement to enter JMET data into DRRS to report readiness capability must be accomplished by hand requiring a significant time investment by the staff. The "Rosetta Stone" provided a means to view the JMETs and assessed values, but there was no relief from the time requirement.

The ability for DRRS via either a middleware solution or a software patch to accept data from a simple EXCEL™ or ACCESS™ spreadsheet would be the most effective solution to the issue. The ease of use of the spreadsheets for data collection, calculation, and even conversion to a JMET hierarchy for DRRS entry has been demonstrated during this research. The Staff's work should be essentially complete when

the exercise or proficiency training is complete, it should not require more time than the exercise entailed to report the results.

What about the ability to permit gross insight into overarching problems encountered in training? Again, the model should permit insights, for this project, into the Doctrine, Organization, Training, Material, Leadership, and Personnel (DOTMLP) framework for resolving the issues encountered. A consistent method of assessing performance, related to specific tasks from the UJTL. These tasks may be in the form of JMETS, Joint Capability Areas (JCAs), critical skills, or some other nomenclature describing some method of accomplishing the necessary tasks to successfully carry out the direction of operational forces. The use of a common set of nomenclature and measures for the performance of those tasks will permit identification of deficiencies or shortfalls, which can be directly related to the DOTMLP framework. Thus, solutions to issues that are generated by a lack of material capability, insufficient numbers of personnel, deficient training, improper organization, or even inadequate doctrine can be identified and implemented with some assurance of success, and perhaps the costs involved can be readily identified as well. This would also constitute a form of ROI in that the costs for correction, the resulting enhancement in operational proficiency, and the assurance that these solutions apply to all similar circumstances can be characterized as ROI.

7.2.3 Extension to National Incident Management Standards

The ability to devise a conversion for JMETs to the appropriate HLS UTL task number for use in the National Incident Management Standards (NIMS) system was discussed previously. Not every JMET that is used in the Noble Resolve series will apply, but a "Rosetta Stone" approach would provide greater visibility to the similarity between the JMETs which address Non-Government agencies such as those addressing Logistics, Politico-Military Support, Consequence Management and CBRNE events. These are common mission sets, as are the requirements to have a comprehensive and useful common operating picture or to share information.

Interagency operations and interoperability have become increasingly more important in the wake of events such as 9/11, Hurricane Katrina, and even the extensive fires in Southern California. The ability for operational level commands to readily join Command and Control architectures or to provide personnel to assist on scene or in Emergency Operations Centers can be enhanced by the continued training for and development of additional JMETs which are the core METs of federal, state, and municipal emergency operations staffs. If these JMETs were readily translatable to the METs within NIMS, then personnel, systems, or support could be identified which would be most suitable for use or to fill unanticipated shortfalls.

7.2.4 Potential Impacts due to Formalized Cost Collection and Assignment Methods

Other than identifying the JMETs or METs which define operational readiness, whether a regime for performance assessment is implemented or not, the critical task in being able to show ROI is developing a formal cost collection methodology. It is insufficient to merely list all resources expended to accomplish the training objectives, this only provides the same macroscopic view of costs which are related in some indefinable way to a list of JMETs.

In order to provide accuracy and accountability to the taxpayers, fully accounting for the resources expended and the JMETs which these were assigned to is necessary. There may be instances where high costs are unavoidable to achieve a specific level of readiness or to train staffs for a very specialized set of missions. That does not mean the expenditures are bad IF the reason for the expenditure and the results of the training can be demonstrated. The costs that are most damaging are those that are unanticipated and then unreported due to embarrassment. Yes, in fact, the assignment of costs to JMETs is Activity Based Costing/Management (ABC/M) but it is more mission oriented. Those costs that do not apply to the core JMETs are still accounted for, but the reason for those expenditures can now be examined and a determination made as to the necessity of the expenditure.

Identifying the categories and JMETs to which costs should be matched at the beginning of the training event is the only effective way to implement the formality

needed to determine the ROI for any event. Demanding formality will slowly overcome the cultural obstacles which currently permit the various segments of any organization to believe that they are either exempt from reporting or that the costs they incur are necessary and need not be justified.

7.3 Recommendations for Further Research

7.3.1 Expansion of Database Use

The late discovery that the database application promised at the beginning of the research was incapable of performing the mathematical functions which were implemented in the EXCEL™ spreadsheets being used for performance assessment was fortunately only disappointing, not disabling.

The logical extension of the methodology is to pair or perhaps migrate the functions developed within EXCEL™ to a more powerful relational database format. The extension to a relational database would permit very quick global changes, sorts, and if development was continued a ready means to link JMETs to UTL METs or Joint Capability Areas.

The other long term need, especially for ROI purposes, is to be able to continually archive data in a manner that cost associations by JMET for all events can be displayed. Overall performance trends which are operational level command neutral as well as scale and environment neutral could be more easily examined to determine common shortfalls and thus implement widespread training or system solutions.

A final example of the use of a database application would be the far more rapid ability to generate the performance assessment tool by using the sort functions for JMETs or by Training Objectives if so desired. This would also permit global changes to Skill Assignment areas when appropriate or desired as well as the ability to easily change weights for measures. The listing discussed herein is by no means exhaustive, but these very simple extensions of the research would significantly enhance the work

accomplished during the research and make the methodology even more adaptable to whatever type of operational level staff or scenario was required.

7.3.1.1 Limitations of Rational Requisite Pro

The primary focus of this section is to recommend further research areas or possibilities, but it would be inappropriate if the shortcomings of the RATIONAL REQUISITEPRO™ software were not addressed.

The two major shortcomings are the inability for the software to accomplish calculations beyond the four basic math functions or some limited derivatives thereof, such as percentages or the simple act of raising a number to a power.

The LDF method proposed, used and refined during this research could not be replicated in RATIONAL REQUISITEPRO™. Separating these calculations into separate worksheets was possible, but would have become ponderous and confusing.

Although it is not uncommon for database applications to require all data to be entered without spaces, the separation of either Skill Assignment areas or OE groupings (and for the researcher inserting brightly colored dividing lines) by spaces was almost unavoidable in order to check the accuracy of the calculations. It is quite honestly much easier to explain to people, even if the computers conducting the calculations do not require the spaces. Unfortunately, the inability to easily export the EXCEL™ spreadsheets used for performance assessment tool to RATIONAL REQUISITEPRO™ and even more disconcerting, import any data from RATIONAL REQUISITEPRO™ into a spreadsheet with all the mathematical functions intact, was the issue that finally made using the software impractical.

A final deficiency, which can probably be overcome, was raised in the Chapter Four discussions regarding multiple measure number assignments for measures which could apply to multiple OE groups, APE functions, or Skill Assignment areas. The redundancy and thus inaccuracy introduced because of this numbering requirement was not fully realized until the staff at Tactical Training Group Atlantic noted the overlaps in

measures and the differences in weights for them. There is probably a routine within RATIONAL REQUISITEPRO™ which can be implemented to avoid this issue and retain the potential to choose the appropriate OE, APE, or Skill Assignment area using only a single measure number, but the personnel assigned to support the research were unaware of such a routine and were never able to overcome the multiple measure assignment issue.

7.3.2 Further refinement of Logarithmic Driving Function (LDF) Approach

Though the ability to accurately and reliably assess the performance of an Operational Level staff was demonstrated by this research, one objective of the research is still not complete. The LDF approach has been improved from the beginning of the research and various approaches for assigning LDF weighting factors have been tested, evaluated and eliminated. The impact of a High importance Skill Assignment area or OE grouping where substandard performance is observed is readily observed, as are the impacts of Medium importance values.

The research has also proved that too many LDF factors render the approach ineffective, but that using the OE groups, which comprised no more than seven factors yielded results that were generally comparable to either averaging method. This was especially true when using only the High or Medium LDF weighted quantities, which was termed the Limited LDF method.

The potential for this method to be the primary performance assessment approach remains to be fulfilled. The reasoning for using this method is compelling since it is very sensitive to variations in the High or Medium factors, it is insensitive to variations in the LOW factors, and tends to show changes in performance in a manner which does replicate human learning. The inability to determine the underlying causes of the inconsistency in results from this method, which primarily appears to stem from the optimal assignment of LDF weighting factors, has been a bit of a disappointment.

The three methods of assigning these factors, all stemming from the JMET driven approach continually reduced the error between both averaging methods and the LDF

method. This was a very elementary optimization attempt which appears to have bounded the issue, but the optimal method for determining LDF weighting factors remains elusive. The application of a some form of linear regression or a much deeper examination of the differences between two of the models which display disparate behaviors for a quantity such as the Assess value for performance would seem to be the most appropriate method of solving the LDF weighting factor dilemma.

If the LDF weighting factor issue could be reliably implemented the OE grouping calculations within the APE functional areas, then extending the solution to the two-level aggregation method would provide the most appropriate way to link OE performance to APE performance. The APE calculations would now serve the purpose of validating the OE performance values if desired, but the ability to use a single data entry approach would be possible, and either view of performance would be available

7.3.3 Further Refinement of Analytical Hierarchy Processing (AHP) Methodology

One of the most pleasant discoveries of this research effort was the discovery and application of the AHP methodology. The inherent issues with obtaining surveys were solved, first by trial and error and by intuition (the use of JMET ratios in place of the factors discussed by Forman (40)) and then by the implementing the JMET ratios devised by the Operational Level staff. The resolution of the instances where there were no measures assigned to a specific OE grouping of an APE functional area was accomplished with satisfactory results. The limited instances where there were large disparities (more than a quantum level) in the number of measures for different OE groupings did not cause mathematically significant variations in the ultimate performance assessment results observed.

The further development of the AHP method so that it can be applied easily to each of the OE groups Skill Assignment areas may lead to the solution of the LDF approach for performance assessment within the OE areas where large numbers of factors exist. This may be solved because the High or Medium factors can be selected with greater confidence and thus the use of the LDF method for performance could become the primary method for performance assessment calculation. The advantages of this

approach were discussed in the previous section, but the solution may reside in the further refinement of the AHP method to determine Skill Assignment area importance.

7.3.4 Further Refining Sensitivity Analysis

Although some discussion of sensitivity analysis was presented in Section 6.8, further work in that area is recommended. The LDF require further refinement to permit the potential of that calculation method to be useful. While the refinement of the LDF factors is occurring, using sensitivity analysis to assure that the factors that appear to provide more accurate results are not merely coincidental must be carried out.

The review of skill assignment areas to arrive at an optimal set, which can apply in not only military, but non DoD and corporate applications needs to be continued.

Finally, continued refinement of the weighting factors, whether those are simple H/M/L or a variation using AHP or some other technique should also be continued.

7.3.5 Continue Pressing the Boundaries of ROI

As Dust (48), Worthen (49), and Conner (50) have discussed, ROI is evolving from the traditional definitions used by the manufacturing industry. The military, and the other service organizations discussed in this research need a method of portraying the ROI achieved. Training and Operational readiness are requirements for all, and what may have been considered Cost Benefit Analysis in the past is incapable of capturing all the factors which must be considered to determine ROI in the environments examined during this research.

What does ROI mean, and what other ROI portrayals are possible are questions which bear investigation and resolution. Simply continuing to extend, by collecting successive event data, the efforts initiated in this research would yield immediate benefit to experiments such as Noble Resolve. The definitive comparison of LVC to synthetic events can be continued with the possibility that the JTFEX as a Operational Level event could be eliminated, and the time could be allocated for Live training that is very specifically tailored to very perishable tactical skills. The robust and very fast paced

nature of simulation could then make the FST-J series the certification event for a CFMCC staff or serve as the basis for true multi-Service Component Commander training. The ability to predict cost avoidance and implement training to reduce costs for follow on events such as FA PANAMAX for Operational Level staffs has implications for budgeting as well as permitting all Operational Level Commanders to have a common basis for requesting training support.

The ability to portray ROI reliably is an enabling capability for all of the above, but as suggested previously, this is not an exhaustive list and there is no intention to limit the potential ROI opportunities that could arise. As long as performance can be assessed, or at least the JMET basis for performance is determined (and the JMETs cited are considered accomplished) and the costs required to accomplish the training are linked to the same JMETs, almost any formulation of ROI desired can be derived.

7.4 It's NOT about Black Hats, it's About Deliberate Choices

The vision and inspiration provided by the Operational Level Commander to his staff and the acceptance of the opportunity to implement the methodology developed during the course of this research was both liberating and refreshing. The responsibility placed on the Commander (or analogously Federal Director, Governor, or Mayor) to successfully direct large scale operations is daunting. The accountability demanded of these same personnel is inconceivable to all but a very few people. His confidence in his abilities and the insights he would gain by using a methodology based in doctrine to assess the processes required to execute JMETs with a numerical method should be the template for future Operational Level exercise assessment.

The basis for performance assessment is NOT to merely or even primarily arrive at some set of numerical values to be used for inane comparison or discussion. It is to determine which processes are most effective in accomplishing the JMETs (METs) which are necessary for mission success. There is no absolute ROI other than that proposed by General Luck (34), but the lessons of Iraq should be motivation enough to seek ways to better prepare our Operational Commanders for the very dynamic and uncertain environments they may face. Unless the Operational Level staffs are prepared

for these situations, the Commanders will spend an inordinate amount of time pursuing solutions to problems that are more appropriate for more junior personnel. The only way to get a staff that can support the Commander so he or she can do the job required is to have a scale neutral, environment neutral, and mathematically reliable assessment capability. Only then can the deficiencies which impair the best possible performance of the staff, augmentation personnel, supporting software, and systems be viewed impartially and objectively and then resolved. The Commander would then have the confidence that the products developed by the staff, the alternatives provided by the staff, and the plans proposed by the staff are the best possible, and the Commander can be free to focus on effectively dealing with a disaster relief mission or defeating a very adaptive and persistent adversary.

Objective self-assessment aided by the unbiased views of experienced outside observers has always been the key to the most successful preparation. This methodology is intended to enhance the self-assessment process. The ability to align costs to that process now creates the basis for compelling arguments for resources, as well as providing a common language for all operational level commanders to articulate their requirements.

Though this researcher would re-iterate the words of Sun Tzu, a more contemporary way of stating what has been developed in this research and which should be pursued to truly determine its limitations comes from the movie "Crimson Tide." As stated by Denzel Washington when there was doubt as to the necessity to launch a nuclear strike, "It's all about knowing." Indeed, it is all about knowing and then acting upon what is known to improve operational readiness.

8 BIBLIOGRAPHY

1. **Joint Chiefs of Staff.** Universal Joint Task List. Washington, D.C., USA : Government Printing Office, 08 01, 2005. CJCSM 3500.4D.

2. **Joint Chiefs of Staff.** Joint Operations. *Joint Publication 3.0.* Washington, D.C., USA : Government Printing Office, 08 01, 2006.

3. **Ambrose, S. E.** *Citizen Soldiers: The U.S. Army from the Normandy Beaches to the Bulge to the Surrender of Germany, June 7, 1944 - June 7, 1945.* New York : Simon & Schuster, 1997. 0684815257.

4. **McKeeby, D.** U.S. Armed Forces Engaged in a Variety of Missions: Security, Counterterrorism and Humanitarian. *CHINFO Clips.* [Newpaper]. Washington , D.C. : Navy Chief of Information, 08 23, 2006.

5. **Suantak, L.** *Modeling, Simulation and Visualization of Stability and Support Operations Using Coevolution: Concepts and Evolution.* Tucson : University of Arizona, 2004. Dissertation. 0-496-90799-9.

6. **United States Congress.** National Defense Authorization Act for Fiscal Year 1987 (Commonly: Goldwater-Nichols Department of Defense Reorganization Act of 1986). Washington : Government Printing Office, 1986.

7. **Joint Chiefs of Staff.** Joint Training Manual for the Armed Froces of the United States. [CJCS Manual]. Washington , D.C. : Government Printing Office, 09 01, 2002. CJSCM 3500.3A.

8. *Measuring the Effectiveness of Simulation-Based Training.* **McGinnis, M.L. and Stone, G.F.III.** [ed.] J.M., Morrice, D.J.,Brunner, D.T.,Swain, J.J. Charnes. New york : ACM Press, 1996. Proceedings of the 28th Conference on Winter Simulation. pp. 934-938. 0-7803-3383-7.

9. **Clark, D.R.** ISD Handbook. [Online] Big Dog, Little Dog, 07 13, 1995. [Cited: 02 20, 2007.] http://www.nwlink.com/~donclark/index.html.

10. **Sloman, M.** *The E-Learning Revolution.* New York : American Management Association, 2002. 0-8144-7155-2.

11. *WARSIM 2000: Combining Multiple Expert Opinions from Subject Matter Experts to Generate Requirements for Staff Traiing at Battalion Level and Above.* **McNett, M.D., Phelan, R.G. Jr and McGinnis, M.L.** Orlando : IEEE, 1997. 1997 IEEE International Conference on Systems, Man, and Cybernetics. pp. 1280-1284.

12. **Kirkpatrick, D. L.** *Another Look at Evaluating Training Porgrams; Fifty Articles from Training & Development and Techinical Training: Magazines Cover the Essentials*

of Evaluation and Return-on-Investment. Alexandria : American Society for Training and Development, 1998. 1562860887.

13. **Jean, G.** Navy's Virtual Training Exercises Expanding in Realism and Scope. *National Defense Industrial News.* 09 2006.

14. **Holden, W.T., et al.** *Developing an Environment for Exploring Distributed Operations: A Wargaming Example.* Department of the Army, U.S. Army Research Institure for the Behavoiral Sciences. Arlington : U.S. ARmy Research Institute for the Behavioral Sciences, 2005.

15. **Cokins, G.** *Activity-Based Cost Management in Government.* Vienna : Mangement Concepts, 2002. 1567261817.

16. **Phillips, J.J.** *Return on INvestment in Training and Performance Improvement Programs.* 2nd. Boston : Butterworth Heinemann, 2003. 0-7506-7601-9.

17. **Kaplan, R.S. and Cooper, R.** *Cost & Effect Using Integrated Cost Systems to Drive Profitability and Performance.* Boston : Harvard Business School Press, 1998. 0875847889.

18. **Luck, Gl.** Insights on Jint Operations: The Art and Science. *A Common Perspective.* 2006, Vols. Volume 14, No.2.

19. **Newell, A and Rosenbloom, P.S.** Mechanisms of Skill Acquisition and the Power-Law of Practice. [ed.] J.R. Anderson. *Cognitive Skills and Their Acquisition.* Hilldale : Lawrence Erlbaum Associates, 1981, 1, pp. 1 - 55.

20. *What Can We Learn from Learning Curves?* **Mayer-Kress, G.J., Newell, K.M. and Liu, Y.** Nashua : New England Complex Systems Institute, 1998. International Conference on Complex Systems.

21. *The Power Law of Practice Repealed: Teh case for an exponential law of practice.* **Heathcote, A. and Mewhort, D.J.K.** 2000, Psychonomic Bulletin and Review, Vol. 7, pp. 185-207.

22. *The Learning Curve.* **Ritter, F.E. and Schooler, L.J.** [ed.] W. Kintch, N. Smelser and P. Baltes. Oxford : Pergamon, 2001, International Encyclopedia of the Social & Behavioral Sciences. 0-08-043076-7.

23. **Startin, W.** Issues Affecting Training Readiness Assessments. [Unpublished Work]. Virginia Beach, VA , USA : s.n., 05 19, 2005.

24. **Tillson, C.F.** *Building a Joint Training Readiness Reporting System.* Institute for Defense Analysis. Alexandria : Institute for Defense Analysis, 1996.

25. **Office of the Secretary of Defense.** *The Defense Readiness Reporting System.* Washington, D.C., USA : OUSD Personnel & Readiness, 03 03, 2006.

26. *Human/Systems Reliability in a Changing Environment.* **Spurgin, A.J.** The Hague : IEEE, 2004. 2004 IEEE International Conference on Systems, Man and Cybernetics. pp. 2746-2751. 10 - 3 October, 2004. 0-7803-8566-7.

27. *Impact of Systematic Data Collection on Training Control Room Crews.* **Spurgin, A.J.** Scottsdale : IEEE, 2002. IEEE 7th Human Factors Meeting. pp. 3-1 to 3-5. 0-7803-7450-9.

28. **Stacy, S.M.** Cores and Competencies. *A History of The Idaho National Engineering and Environmental Laboratory, 1949- 1999.* Idaho Falls : Idaho National Laboratory, 2006, 10, pp. 86-93.

29. **Trion, R.** Computer Simulations Bolstert Joint-Service Combat Training. *National Defense Magazine.* 04 2004. http://www.nationaldefensemagazine.org/issues/2004/Apr/Computer_Simulations.htm.

30. **Kennedy, H.** USJFCOM Steps Up Efforts to Organize Combined Ops. *National Defense Magazine.* January 2003. http://www.nationaldefensemagazine.org/issues/2003/Jan/USJFCOM.htm.

31. **Rodney, D.** Aggregate Performance Measures. [Commander Third Fleet Paper]. San Diego, CA, USA : Commander Third Fleet, 05 07, 2003.

32. **Brown, D.K.** NMETLS and Improving Navy Mission Performance. [Unpublished Paper]. Noroflk, VA, USA : s.n., 05 11, 2005.

33. **Group, Robert Francis.** Robert Francis Group IT Agenda. *RFG Online.* [Online] The Robert Francis Group, 05 21, 2001. [Cited: 09 27, 2006.] http://www.rfgonline.com/subsforum/archive/daily/043001/050201nt.html.

34. **Luck, Gen. Gary.** *Return-on-Investment for Training.* [interv.] M.C. Nesselrode. Suffolk, 10 17, 2006.

35. **Joint Chiefs of Staff.** Joint Task Force Headquarters Master Training Guide. [CJCS Manual]. Washington, D.C., USA : Government Printing Office, 09 01, 2003. CHCSM 3500.05A.

36. **Williams, K., et al.** *An Evaluations of a Mehodology for Cognitively Structuring and Adaptively Sequencing Exercise Content for Embedded Training.* Institute for Simulation and Training, University of Central Florida. Orlando : University of Central Florida, 1989.

37. *Integrated Performance Measurement and Assessment in Distriubted Mission Operations Environments: Relating Measures to Competencies.* **Carolan, T., et al.**

Orlando : National Defense Industrial Association, 2003. Porceedings of the 25th Indtersrevice/Industry Training, Simulation & Education Conference.

38. **Bennett, W.** *Mission Essential Competencies (MEC) for Joint Force Air Component Commander (JFACC).* U.S. Air Force, Air Force Reseach Laboratory. s.l. : Air Force Research Laboratory, 2006.

39. **Standing Joint Task Force Headquarters.** Joint Task Force Headquarters: A Concept of Operations. [Draft Concept of Operations Manual (V2.5)]. Norfolk, VA, USA : US Joint Forces Command; Standing Joint Task Force Headquarters, 10 23, 2006.

40. **Forman, E.H. and Selly, M.** The Analytic Hierarchy Process and Expert Choice. *Decision by Objectives (How to Convince Others You Are Right).* River Edge : World Scientific, 2007, Four, pp. 43 - 127. http://mdm.gwu.edu/forman/DBO.pdf; Accessed February 10, 2007.

41. **Oliver, L.** *Designing Strategic Cost Systems: How to Unleash the Power of Cost Information.* Hoboken : John Wiley & Sons, 2004. p. 220. 0471653586.

42. *Costing Training: A Decision Maker's Dilemma.* **Wright, P.C. and Belcourt, M.** 2, London : MCB University Pressed Limited, 1995, Mangement Decision, Vol. 33, pp. 5 - 15.
http://proquest.umi.com.proxy.lib.odu.edu/login?COPT=REJTPTMmU01EPTQmSU5UPTAmVkVSPTI=&clientId=3505.

43. **Schein, E.** What Smart Trainers Should Know About Organizational Culture. [ed.] L.L. Ukens. *What Smart Trainers Know: The Scerets of Succcess From the World's Foremost Experts.* San Francisco : Jossey-Bass/Pfeiffer, 2001, Six, pp. 54-60.

44. **Kraiger, K.** Decision-Based Evaluation. *Creating, Implementing, and Magaging Effective Training Development.* First. San Francisco : Jossey-Bass Inc., 2002, Eleven, pp. 331-375.

45. **Nelson, M.K. and Bely, D.C.** *Cost-Effective Live-Fire Test and Evaluation Startegies: The Missions and Means Framework.* US Army Material Systems Analysis Activity. Aberdeen : US Army Material Systems Analysis Activity, 2005. Final Report. Delivery Order 196, Contract No. DAAD19-02-D-001.

46. **Fazal, M.B.** *Effectiveness and Cost of Computer Based Instruction for Higher Education and Defense Training: A Meta-Analysis.* University of Southern California. Los Angeles : University of Southern California, 1996. Dissertation. http://proquest.umi.com/pqdweb?did=739421031&sid=6&Fmt=2&clientId=3505&RQT=309&VName=PQD . 0591117681.

47. **Cox, A.J.** *Total Quality Leadership: The Role of Total Quality Leadership in Joint Military Operations.* U.S. Naval War College. Newport : U.S. Naval War College, 1997. pp. 1-18, Student Project. http://stinet.dtic.mil/cgi-

bin/GetTRDoc?AD=ADA325243&Location=U2&doc=GetTRDoc.pdf, Accessed 11/12/2007.

48. **Dust, B.** *The Myth of Training ROI.* Richmond, VA, USA : Media Tec Publishing Inc., Chief Learning Officer Magazine, January 2004. Gyrus Learning Management System Weblog. http://www.gyrus.com/documents/WhitePapers/The%20Myth%20of%20Training%20ROI.pdf.

49. **Worthen, B.** Measuring the ROI of Training. *CIO Magazine.* 02 16, 2001. http://www.cio.com/archive/021501/roi.html, Accessed 6/12/2006.

50. **Conner, M.L.** *How Do I Measure Return on Investment (ROI) for My Learning Program.* Mechanicsville, Va , USA : Learnativity, April 05, 2002. Learnativity.com. http://learnativity.com/roi-learning.html ; Accessed 2/13/2006.

51. *A Business Case for Modeling and Simulation.* **Carter, J.R.** [ed.] W. Waite. Orlando : Simulation Councils Inc., 2001. Summer Simulation Conference. pp. 587-594.

52. *Uses and Benefits for the Validated Stinger Hardware in-the-loop Simulation.* **Olden, H., et al.** [ed.] A Obaidat, A. Nisanci and B. Sadoun. Chicago : Simulation Councils Inc., 1999. Summer Simulation Conference. pp. 555-559. http://www.scs.org/getDoc.cfm?id=419.

53. *Determining the Value of Simulation.* **Gordon, S.C.** [ed.] W. Waite. Vancouver : Simulation Councils Inc., 2000. Proceedings of the 2000 Summer Simulation Conference. pp. 969-973. http://www.scs.org/getDoc.cfm?id=1126#search=%22determining%20the%20value%20of%20simulation%22.

54. **Nesselrode, M.C.** *Simulation of s Surface Combatant Maneuvering in Four Degrees of Freedom.* U.S. Naval Postgraduate School. Monterey : U.S. Naval Postgraduate School, 1985. Thesis.

55. **McGibbon, T.** *A Business Case for Software Process Improvment Revisited.* Air Force Research Laboratory/Information Directorate. Rome, NY : Data & Analysis Center for Software, ITT Industries, Advanced Engineering and Sciences Division, 1999. Revision to previous State of the Art Report. http://www.dacs.dtic.mil/techs/roispi2/index.html.

56. **Cross, J.** *A Fresh Look at ROI.* [Article] [ed.] R. Ellis. Alexandria, VA, USA : Learning Circuits, American Soecitety for Training & Development, January 2001. http://www.learningcircuits.org/2001/jan2001/cross.html , Accessed January 22, 2007.

57. **Quinn, D. J.** *The Goldwater-Nichols DOD Reorganization Act: A Ten-Year Retrospective.* Washington : National Defense University Press, 1999. p. 77. 1579060072.

58. **Wess, R.** *Leadership Secrets of Attila the Hun.* New York : Warner Books, 1989. p. 110. 0446515167.

59. **Axelrod, A.** *Patton on Leadership: Strategic Lessons for Corporate Warfare.* Paramaus : Prentice Hall Press, 1999. 0-7352-0297-4.

60. **Cohen, W.A.** *Wisdom of the Generals; How to Triumph in Business and in Life.* Paramus : Prentice Hall Press, 2001. 0-7352-0167-6.

61. *Wargame-Based Training Systems.* **Ogus, R.L., Jancaitus, J.R. and Hauck, R.E.** [ed.] T.I Oren, E.M. Delfosse and C.M. Shub. Piscataway : IEEE Press, 1981. Proceedings of the 13th Converence on Winter Simulation. Vol. 2, pp. 667-677. http://delivery.acm.org.proxy.lib.odu.edu/10.1145/810000/802647/p667-ogus.pdf?key1=802647&key2=5452500511&coll=ACM&dl=ACM&CFID=78150863&CFTOKEN=50128118.

62. **Department of Defense.** Modeling and Simulation Master Plan. DoD 5000.59-P [ed.] Under Secretary of Defense for Acquistion and Technology. *DoD 5000.59-P.* [DOD Directive]. Washington , D.C., USA : Government Printing Office, October 1995.

63. **IBM Corporation.** IBM RATIONAL REQUISITEPRO. Somers, N.Y., USA : IBM Corporation Software Group, 2006.

64. **Department of Homeland Security.** Homeland Security Universal Task List. [Homeland Security Directive]. Washington, D.C., USA : Department of Homeland Security, September 15, 2007.

65. **Tzu, Sun.** *The Art of War: The New Translation.* [ed.] J.H. Huang. New York : William Morrow and Company, Inc., 1993. 0-688-12400-3.

66. **Commander Southern Command.** CJTF South Training Plan. [Commander Southern Command Instruction]. Miami, Fl, USA : Commander Southern Command, March 15, 2007.

67. **Joint Chiefs of Staff.** Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms. [CJCS Publication]. Washington, D.C., USA : Government Printing Office, April 12, 2001. http://www.dtic.mil/doctrine/jel/new_pubs/jp1_02.pdf; Accessed 10/15/2007.

68. **Commander Second Fleet.** Maritime Headquarters with Maritime Operations Center Concept of Operations. [Concept of Operations]. Norfolk, VA, USA : Commander Second Fleet, March 13, 2007. p. 128.

69. **Startin, Dr. William.** *Assessment Methodology Theory.* [interv.] M.C. Nesselrode. Virginia Beach, July 19, 2007.

70. **Joint Forces Command (J-9 Directorate).** Final After Action Review for Noble Resolve 07-2. [Slide Presentation]. Suffolk, VA, USA : Joint Forces Command (J-9 Directorate), August 24, 2007.

71. **Joint Forces Command (J-9 Directorate).** Noble Resolve 07-2 Final Draft Report. Suffolk, VA, USA : Joint Forces Command (J-9 Directorate), October 29, 2007.

72. **Chanik, VADM M. E.** *FA PANAMAX Exercise Assessment Goals.* [interv.] M. Nesselrode. Norfolk, December 17, 2006.

73. **Flynn, S.E.** *The Edge of Disaster: Rebuilding a Resilient Nation.* New York : Random House, Inc., 2007. 978-1-4000-6551-6.

74. **Chanik, VADM M.E.** *Briefing of FA PANAMAX 07 Results.* [interv.] M.C. Nesselrode. Norfolk, Septermber 25, 2007.

75. **Startin, Dr. William.** *Further Discussion Regarding Assessment Methodology.* [interv.] M.C. Nesselrode. Suffolk, September 26, 2007.

76. **Morris, W., [ed.].** *the American Heritage Dictionary of the English Language.* New York : American Heritage Publishing Co., Inc., 1971. 395-09064-4.

Appendix A – Sample Rosetta Stone

C2PA	Observer Name	Description of task	OPR	Attribute	Observed Value	Proposed Standard(T2)	T1	T3	T4
OPP		Establish Operational Protection Working Group (OPWG)							
55021		Protective supporting schemes established within operational	FOPS	Pct		90-95	>95	80-89	<80
55022		Establish OPWG within 2 hours of OPT activation	FOPS	Hours		2	<2	3-4	>4
55023		OPWG composed of OPP staff and representatives from each component, COCOM, interagency, and appropriate JPG representatives	FOPS	Pct		80-89	>90	70-79	<70
55024		Component plans assessed for risk	FOPS	Y/N		Y(.85)		N(.50)	
FOPS		Expand Future Ops (FOPS) Cell							
55081		Planners, functional representation, deployment planner, JOPES operator, TPFDD analyst, component LNOs, COCOM LNO, SJFHQ augmentation, interagency and stakeholder representatives as available)	FOPS	Pct		80-89	>90	70-79	<70
55082		Establish collaborative planning environment that facilitates efficient information exchange between JTF distributed staff, COCOM, and subordinate commands	FOPS	Hours		18-24	<18	24-36	>36
55083		Identify system requirements for communications with outside stakeholders	FOPS	Hours		18-24	<18	24-36	>36
55084		Establish Red and Blue cell capability and expertise in each OPT	FOPS	Y/N		Y(.85)		N(.50)	
55085		Identify FOP representation at JTF SOP-defined boards and cells	FOPS	Y/N		Y(.85)		N(.50)	
FOPS, Fplans		Conduct Mission Analysis							
55245		Prepare WARNORD for issuance by COPS	FOPS	Hours		4-6	<4	6-8	>8
FOPS, Future Plans		Develop COAs for analysis, comparison, and selection							
55262		FOPS develops COAs for any missions that would occur outside of 24 hours but within current phase (branch plans)	FOPS	Hours		18-24	<18	24-36	>36
55263		Comparison criteria is identified, and COAs are compared according to established comparison	FOPS	Y/N		Y(.85)		N(.50)	

Appendix B – Sample Observer Sheet

C2F#	Observer Name	Description of task	OPR	Attribute	Observed Value	Proposed Standard(T2)	T1	T3	T4
OFP		Establish Operational Protection Working Group (OPWG)							
55021		Protective supporting schemes established within operationa	FOPS	Pct		90-95	>95	80-89	<80
55022		Establish OPWG within 2 hours of OPT activation	FOPS	Hours		2	<2	3-4	>4
55023		OPWG composed of OFP staff and representatives from each component, COCOM, interagency, and appropriate JPG representatives	FOPS	Pct		80-89	>90	70-79	<70
55024		Component plans assessed for risk	FOPS	Y/N		Y(.85)		N(.50)	
FOPS		Expand Future Ops (FOPS) Cell							
55081		Planners, functional representation, deployment planner, JOPES operator, TPFDD analyst, component LNOs, COCOM LNO, SJFHQ augmentation, interagency and stakeholder representatives as available)	FOPS	Pct		80-89	>90	70-79	<70
55082		Establish collaborative planning environment that facilitates efficient information exchange between JTF distributed staff, COCOM, and subordinate commands	FOPS	Hours		18-24	<18	24-36	>36
55083		Identify system requirements for communications with outside stakeholders	FOPS	Hours		18-24	<18	24-36	>36
55084		Establish Red and Blue cell capability and expertise in each OPT	FOPS	Y/N		Y(.85)		N(.50)	
55085		Identify FOP representation at JTF SOP-defined boards and cells	FOPS	Y/N		Y(.85)		N(.50)	
FOPS,Fplans		Conduct Mission Analysis							
55245		Prepare WARNORD for issuance by COPS	FOPS	Hours		4-6	<4	6-8	>8
FOPS, Future Plans		Develop COAs for analysis, comparison, and selection							
55261		FOPS develops COAs for any missions that would occur outside of 24 hours but within current phase (branch plans)	FOPS	Hours		18-24	<18	24-36	>36
55265		Comparison criteria is identified, and COAs are compared according to established comparison	FOPS	Y/N		Y(.85)		N(.50)	

Appendix C -- Skill Assignment Area Definitions

The Skill Assignment Area definitions supplied below are shown with the definition of the word as found in The American Heritage Dictionary (76) (if applicable). Further modification to show the usage by the researcher is then displayed.

Acceptability – Satisfactory; adequate.

Accountability – Answerable. Capable of being explained.

Accuracy – Exactness; correctness.

Adaptability – Capable of adapting or of being adapted.

Agility – The state or quality of being agile; nimbleness; briskness.

Authority – The right and power to command, enforce laws, exact obedience, determine, or judge.

Balance – To reach or achieve a state or position between extremes. A Staff's ability to account for all inputs or factors and reach a logical end state which optimizes the needs of the Commander.

Clear Communications – The ability to relay orders and plans in an unambiguous fashion so that Higher Headquarters, Horizontal components, or subordinate commands can respond to the orders or plans with, at most, minimal clarification.

Competency – The state or quality of being capable or competent; skill; ability.

Confidence – Trust in a person or thing.

Courage – The state or quality of mind or spirit that enables one to face danger with self-possession, confidence, and resolutions; bravery; valor. Courage suggests a reserve of moral strength on which one may draw in time of emergency. The ability of the Commander to be resolute despite potential difficulty or uncertainty.

Decisiveness – Having the power to settle a dispute or doubt; conclusive. Characterized by decision and firmness, resolute; determined. Beyond doubt; unmistakable; unquestionable.

Efficiency – The quality or property of being efficient. The degree to which this quality is exercised.

Flexibility – Responsive to change.

Initiative – The power, ability, or instinct to begin or to follow through with a plan or task; enterprise and determination.

Leadership – The capacity to be a leader; ability to lead.

Liaison – An instance or means of communication between bodies, groups, or units.

Negotiation – The act or procedure of negotiating.

Objectivity – The state, condition, or quality of being objective.

Perceptiveness – Having the ability to perceive; keen in discernment; knowing. Marked by discernment and understanding; sensitive.

Preparedness – The state of being prepared; especially military readiness for war.

Projection – A plan for an anticipated course of action.

Responsibility – A thing or person that one is answerable for; a duty, obligation, or burden.

Situational Awareness – The ability to understand and comprehend the current operational situation using all available inputs including data, voice communications and displays.

Suitability – Appropriate to a given purpose or occasion.

Thoroughness – Painstakingly accurate or careful.

Timeliness – Occurring at a suitable or opportune time; well-timed.

Trust – Firm reliance on the integrity, ability, or character of a person or thing; confident belief; faith.

Appendix D – Noble Resolve 07-2

Day 2 AAR

Comments for Day 2

1. **Oregon Nat'l Guard –**
    a. Military imagery support
        i. Useful, goes through FEMA*
        ii. Aids in Command decision*
    b. Need JIEE Multiple user view
        i. Lack of SA significant*
    c. ONG Requirements
        i. Use for soldiers for aid*
        ii. Operationally hampered without this aid
    d. iCAV Tech issues
        i. Without end user access*
        ii. Extreme SA and information processing impacts
2. **Oregon Fusion Center**
    a. iCAV
        i. Useful if funded, better access and information

b. Titan needs to look at Criminal activity and needs the intell support to combat

    i. Much busier and more critical during an event or real situation where these elements could play*

c. **RFI process**

    i. Need to get this comment—Intell is big, regardless of the kind of situation or scenario that is being dealt with. There seems to be a lack of response by JTF HD on RFIs for various agencies, notably Oregon Fusion Center.*

3. Nat'l Guard Bureau

a. Connectivity

    i. NG internal communications interoperability were validated, but need more procedures.*

    ii. ICAV not useable within or without NG

    iii. Due to the above, have SA and procedural requirements that need to be addressed

    iv. Also now understand the pre-planning and interoperability requirements to start at a more fully operable state*

b. Lack of ICAV operability and use for COP is an issue

    i. SA very significantly degraded and cannot assist in COP or SA build for other agencies.*

4. Joint Task Force Homeland Defense

a. Inability to access user generated map/modeling/real time COP product on ICAV prohibited good SA for JTF*

    i. Attempted to generate a near real time COP via iCAV but could not, thus had very limited incident visibility from subordinate units.

5. Task Force Guam

a. iCAV as COP

i. Guam doesn't have specific geography in iCAV*

ii. Critical infrastructure limitations not being available limit usefulness of iCAV*

iii. There is potential, but it isn't there yet

iv. Can only get data to HHQ via phone and emails*

v. Database alteration or management is a major issue that needs to be addressed, or at least have training available vice the imbedded tools.*

6. ROK Response Cell

a. Enhanced connectivity tools iCAV, IWS, ENN, JOT, and ENG are not available to the ROK response cell due to tech issues.*

b. Firewalls on each side prevent connection*

c. CDCIE and SIFG are limited alternatives

d. Limited SA impairs quality and speed of response

e. Traditional means such as telephones can be useful, but provide minimal communications.

7. NORTHCOM

a. NC, DHS, DoD, NGB all have collaboration needs*

i. HSIN expanded use may enhance collaboration

ii. Need to have DoD and NGB input TTPs for infusion in the database for iCAV

iii. Synchronization can only occur for NG-DSCA if the data is timely and accurately populated and if there is reciprocity between HSIN/iCAV and DHS data.*

8. Interagency

a. Conference with USAID, USAF, FEMA, as well as others, permitted a look at streamlining a new US method for getting assistance from non DoD and especially non-US agencies.*

b. Also very different when looking at NON mil-mil assistance requests

c. Transportation and personnel transfer issues were key in the discussion.

d. Looking to see how these can work in a table top

Appendix E – Noble Resolve 07-2

Daily After Action Review Slides

Slide 1

NR07-2
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 2

NR07-2
Oregon National Guard
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 3

Oregon National Guard
(NR07-2, 21 Aug)

Insight #1 – Military Imagery Support
Insight #2 – JIEE Multiple User Viewing
Insight #3 – ONG Requirements for Execution
Insight #4 – Technical Issues with iCAV

Slide 4

Military Imagery Support

- Observation and/or discussion
  - Education on process – request to FEMA then DCO
- Operational Impacts
  - Improve ground view of area for command decisions

Slide 5

JIEE Multiple User Viewing

- Observation and/or discussion
  - Viewing other state/territory inputs.
- Operational Impacts
  - Maintain situational awareness from each perspective through shared communication device.

Slide 6

ONG Requirements for Execution

- Observation and/or discussion
  - Reviewing processes required to take next steps toward supporting.
- Operational Impacts
  - Inability to proceed without declarations

Slide 7

Technical Issues With iCAV

- Observation and/or discussion
  - Pre-loaded maps limited user input
  - Inability to utilize address function
  - End user needs maximum privileges
- Operational Impacts
  - Time delays in presenting information accurately

Slide 8

NR07-2
Oregon National Guard
Daily AAR Process

Comments/Questions?

Slide 9

NR07-2
Titan Fusion Center
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 10

TITAN Fusion Center
(NR07-2, 21 August)

- Insight #1: iCAV appears to be a useful tool (with enough bandwidth and DHS funding)
- Insight #2: Improved our RFI Process
- Insight #3: TITAN would have more involvement in LE Intel support for Portland disaster

Slide 11

TITAN Fusion Center
Insight #1: iCAV appears to be a useful tool (with enough bandwidth and DHS funding)

- Observations and/or discussion points:
  - iCAV was useful in tracking damage in Portland, as well as maritime picture
  - Would consider use if DHS where to fund system and architecture.
- Operational Impact:
  - iCAV could enhance overall Situational Awareness

Slide 12

TITAN Fusion Center

Insight #2: Improved our internal RFI Process

- Observations and/or discussion points:
  - RFI Process today was centrally managed
  - Will review internal process for future events
- Operational Impact:
  - Quicker delivery and response during future NR events this week.

Slide 13

TITAN Fusion Center

Insight #3: TITAN would have more involvement in LE Intel support for Portland disaster

- Observations and/or discussion points:
  - Given potential criminal impact during Portland disaster, TITAN would likely have been tasked with greater intel support
- Operational Impact:
  - Much busier during an event of this magnitude

Slide 14

NR07-2
Titan Fusion Center
Daily AAR Process

Comments/Questions?

Slide 15

NR07-2
National Guard Bureau
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 16

Slide 17

Slide 18

National Guard Bureau

Insight # 2: Effectiveness of COP (iCAV)

- Observation:
  - iCAV has served as a common data base, it has not allowed us to share data or products developed by the users
  - Have not been able to determine information requirements to support the DHS and USNORTHCOM COP systems?
  - What are the interagency COP requirements?
- Operational Impact:
  - Impact on NGB SA
  - Unable to contribute to DHS & USNORTHCOM SA

Slide 19

NR07-2
National Guard Bureau
Daily AAR Process

Comments/Questions?

Slide 20

NR07-2
JTF-HD
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 21

JTF-HD

- Insight # 1: *Inability to access user generated map/modeling/real time COP product on iCAV*

Slide 22

Joint Task Force Homeland Defense

Insight #1: *Inability to access user generated map/modeling/real time COP product on iCAV*

- Observation and/or discussion points:
  - TF Guam attempted to generate a real time COP via iCAV for but could not. GIS generated product may not have been e-mailable due to size.
- Operational Impacts:
  - Limited incident visibility from sub-ordinate unit.

Slide 23

NR07-2
JTF-HD
Daily AAR Process

Comments/Questions?

Slide 24

NR07-2
Task Force Guam
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 25

Task Force Guam

- Insight # 1: Use of iCAV as a COP
- Insight # 2: iCAV as a Data Sharing Tool

Slide 26

Task Force Guam

Insight # 1: Limited COP when data is unavailable

- Observation and/or discussion points:
  - There is no specific data loaded into iCAV for Guam (unlike data loaded for the continental U.S.)
  - Critical infrastructure data i.e roads, power plants, and points of interest (hospitals, fire departments, police, etc) are not available on iCAV
- Operational Impact:
  - Without better defined data loaded into iCAV, the use of this tool as a Common Operations Picture (COP) was ineffective
  - Suggest infrastructure and imagery data for Guam be loaded into iCAV before this tool is offered as an info sharing/situational awareness tool

Slide 27

Task Force Guam

Insight # 2: iCAV as on-the-fly COP

- Observation and/or discussion points:
  - iCAV training demonstrated how a "snapshot" could be emailed to others but that's a static COP which other systems (even MS Powerpoint) can accomplish
  - Experiment/training does not demonstrate how (or if) iCAV can create a "dynamic COP" that one could push to other users (i.e. HHQ) information as it changes
- Operational Impact:
  - iCAV has potential to be an effective tool for sharing data.
  - iCAV as tool to develop and share situational awareness and between Guam and JTF-HD is limited to "static display" only
  - Dynamic updates occur via phone calls and emails

Slide 28

NR07-2
Task Force Guam
Daily AAR Process

Comments/Questions?

Slide 29

NR07-2
ROK
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 30

ROK Response Cell

(NR07-2, 21 August)

- Title: Enhanced Connectivity

Slide 31

Title: Enhanced Connectivity

- Observations and/or discussion points:
  - iCAV, IWS, ENN, JOT, and ENG are experimental tools & technology not available to the ROK Response Cell.
  - J9 JFCOM & ROK each have firewalls that prevent us from talking to each other.
  - CDCIE & SIFP are limited alternate options.
- Operational Impact:
  - Limited situational awareness impairs quality and speed of response.
  - Must use traditional means of the telephone to ensure minimal communication.

Slide 32

NR07-2
ROK
Daily AAR Process

Comments/Questions?

Slide 33

NR07-2
USNORTHCOM
Daily AAR Process

Tuesday 21 AUG 07
Overview

Slide 34

USNORTHCOM

- Insight # 1: NC & DHS, DoD, NGB Collaboration Needs

Slide 35

Title: NC & DHS, DoD, NGB Collaboration Needs

- Discussion points:
  - N-NC staff advocating the use of HSIN and recommendations regarding limited DoD administrative and/or sponsorship rights to selected sites
  - DoD and NGB input TTPs for infusion of DoD/NGB data to iCAV (with future database replication between iCAV and the USNORTHCOM SAGE – Situational Awareness Geospatial Enterprise)
- Operational Impact: Intent to synchronize – NC DSCA data is timely and accurately populated in HSIN/ICAV with reciprocity of DHS data on SAGE

Slide 36

NR07-2
USNORTHCOM
Daily AAR Process

Comments/Questions?

Slide 37

NR07-2
Interagency
Daily AAR Process

Tuesday 21 AUG 07
Overview

Appendix F – COCOM JMETL for FA PANAMAX Execution

1. Purpose. To present a summary of joint tasks the Commander considers essential to accomplish assigned missions.

2. Description. This matrix identifies the command's mission capability requirements for the core JTF HQ, and for those necessary by discrete mission requirements, and forms the basis for the command's joint training program (requirements-based training). The complete JMETL, including JMETs, responsible organization, conditions, and standards can be found by referring to Tab C of this Joint Training Plan.

CJTF South Joint Mission Essential Tasks

Executive Core and Mission-Discrete Task Summary

Mission Essential Tasks (Office of Primary Responsibility - OPR)		CORE	6075	6120	6150	6175	6225	6333	6400	6601
OP 1.1	Conduct Operational Movement (Op/Plans-COPS)	X		X	X	X	X	X	X	
OP 1.2	Conduct Operational Maneuver and Force Positioning (Op/Plans-COPS)		X							
OP 1.4.4	Conduct Maritime Interception (Op/Plans-COPS)			X						
OP 1.5	Control Operationally Significant Areas (Op/Plans-COPS)								X	X
OP 3.1	Conduct Joint Force Targeting (Op/Plans-Fires)		X				X	X		
OP 4	Provide Operational Logistics and Personnel Support (OSI)	X			X	X	X	X	X	X
OP 4.4	Coordinate Support for Forces in the Joint Operations Area (OSI)			X						
OP 4.7	Provide Politico-Military Support to Other Nations, Groups and Government Agencies (Op/Plans-IA)	X	X	X						
OP 5.2	Assess the Operational Situation (ONA-JAC)	X	X		X	X	X	X	X	X
OP 5.5	Establish, Organize, and Operate a Joint Force Headquarters (Op/Plans-Training)	X	X	X	X	X	X	X	X	X
OP 5.7	Coordinate and Integrate Joint/Multinational and Interagency Support (Op/Plans-IA)	X	X		X	X	X	X	X	X
OP 6.2	Provide Protection for Operational Forces, Means, and Noncombatants (OFP)	X		X	X	X	X	X		
OP 6.5	Provide Security for Operational Forces and Means (OFP)									X

(S) CAMPLAN 6075 – (Classified)
(S) FUNCPLAN 6120 – (Classified)
(U) FUNCPLAN 6150 – Humanitarian Assistance / Foreign Disaster Relief
(U) FUNCPLAN 6175 – Defense Support of Civil Authorities
(S) FUNCPLAN 6225 – (Classified)
(S) CONPLAN 6333 – (Classified)
(S) CONPLAN 6400 – (Classified)
(S) CONPLAN 6601 – (Classified)

This page intentionally left blank

Appendix G – FA PANAMAX Costs

FA PANAMAX Fixed and Direct Costs	Synthetic	JMET
RearCMDE Connectivity	$ 40,000.00	OP 5.5
Security Materials	$ 600.00	OP 5.5
Services/ Materials	$ 8,249.65	OP 5.5
Services/ Materials	$ 2,378.80	OP 5.5
Hazmat Services	$ 649.48	OP 5.5
Hazmat Services	$ 740.28	OP 5.5
Vehicle Rental	$ 51,000.00	OP 5.5
Transport	$ 4,500.75	OP 5.5
	$ 108,118.96	OP 5.5

Note: 1) Services, materials, transport, and rental were all connected with the movement, setup, and retrograde of the DJC2 system
2) SOUTCOM and JPO DJC2 program costs are not reflected in either direct or SOUTCOM one-time/event specific totals

Appendix G – FA PANAMAX Costs (continued)

FA PANAMAX Personnel Costs		Synthetic					
	Exercise Labor						
	Apportioned Contract Labor Services		$ 41,377.75	OP 5.5		OP 1.1	$ 269,690.30
	Knowledge Mgr. JOPES, LOG, FDO		$ 58,070.85	OP 4.0			
						OP 3.1	$ 269,690.30
	CERT Lodging		$ 340.00	ACB			
	Cert TAD/Food		$ 148.80	ACB		OP 4	$ 316,146.98
	Cert TAD Funding		$ 4,000.00	ACB			
	Cert TAD Logding/Transport		$ 130,500.00	ACB		OP 5.2	$ 269,690.30
	Cert TAD Logding/Transport		$ 1,425.00	ACB			
	Interpreter		$ 1,552.50	OP 5.7		OP 5.5	$ 311,068.05
	Lodging		$ 1,254.30	ACB			
						OP 5.7	$ 282,856.97
	Civilian Labor Costs		$ 168,428.00	ACB			
	JST Labor Costs		$ 1,258,045.00	ACB		OP 6.2	$ 269,690.30
	JST Travel		$ 188,446.00	ACB			
	JST Contract Costs		$ 18,785.00	ACB			
	Government Travel		$ 116,460.00	ACB			
			$ 1,988,833.20				$ 1,988,833.20

Appendix G – FA PANAMAX Costs (continued)

FA PANAMAX Indirect Costs		Synthetic	
	CIS Work station Materials	$ 16,000.00	OP 5.5
	VTC Repairs	$ 500.00	OP 5.5
	WSC - 3 Repairs	$ 13,000.00	OP 5.5
	Document Scanners	$ 40,000.00	OP 5.5
	Adobe Acrobat Pro	$ 7,500.00	OP 5.5
	Indirect Totals	$ 77,000.00	OP 5.5

Appendix G – FA PANAMAX Costs (continued)

FA PANAMAX One Time or Event Specific	Synthetic	
Extenal M&S Support		
Hardware and Software Support	$20,601.37	OP 5.5
Uniform Items	$ 6,793.28	OP 5.5
Instructional Materials/Maps	$ 8,250.08	OP 5.2
Presentation Materials	$13,688.06	OP 5.2
Signs	$ 500.00	Uniform ACB
	$49,832.79	

Appendix G – FA PANAMAX Costs (continued)

FA PANAMAX One Time or Event Specific	Synthetic	
Extenal M&S Support		
Hardware and Software Support	$20,601.37	OP 5.5
Uniform Items	$ 6,793.28	OP 5.5
Instructional Materials/Maps	$ 8,250.08	OP 5.2
Presentation Materials	$13,688.06	OP 5.2
Signs	$ 500.00	Uniform ACB
	$49,832.79	

Appendix G – FA PANAMAX Costs (continued)

	Overall Cost Breakdowns			
Direct		$ 108,118.96		OP 5.5
Personnel		$ 269,690.30		OP 1.1
		$ 269,690.30		OP 3.1
		$ 316,146.98		OP 4
		$ 269,690.30		OP 5.2
		$ 311,068.05		OP 5.5
		$ 282,856.97		OP 5.7
		$ 269,690.30		OP 6.2
Indirect		$ 77,000.00		OP 5.5

Appendix G – FA PANAMAX Costs (continued)

One Time/Event Specific		$ 27,466.08		OP 5.5
		$ 22,009.57		OP 5.2
		$ 71.43		OP 1.1
		$ 71.43		OP 3.1
		$ 71.43		OP 4
		$ 71.43		OP 5.7
		$ 71.43		OP 6.2
		$ 2,223,784.95		

Appendix G – FA PANAMAX Costs (continued)

JMET Breakdown		
OP 1.1		0.1213
OP 3.1		0.1213
OP 4		0.1422
OP 5.2		0.1312
OP 5.5		0.2355
OP 5.7		0.1272
OP 6.2		0.1213
		1.0000

Appendix H – FY06 MCO Live and Synthetic Costs

Fixed and Direct Costs	MCO (JTFEX 06) Direct Costs			Live						Synthetic	
Port Services				$ 110,607.98	1.2					$ 731,585.90	
		Dep/Arr		$ 34,939.00	1.2			Hotel Svcs		$ 731,585.90	1.2
				$ 34,939.00	1.2						
		PW Connect		$ 20,364.99	1.2						
		PW Disconnect		$ 20,364.99	1.2						
Fuel				$ 5,791,458.96						$ -	
		CVN		$ -							
		LHD		$ 157,278.24	3.2						
		CG		$ 207,658.08	6.1						
		DDG		$ 355,093.20	6.1						
		LSD		$ 40,748.40	3.2						
		LPD		$ 66,573.36	3.2						
NOTES: 1) Burn rates by class, F76/DFM at $2.52/Gallon --> 42 Gallons = 1 Barrel, Based upon 16kt SOA, costs per day											
t $73,158.59 per day (mar 2006)											
ta from FST 06-2 and JTFEX 06-1								JTEN Costs		$ 153,438.74	3.2/6.1

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

Personnel Costs		Live		MCO FST-J 06 Personnel Costs		Synthetic
Training Support						
	Mobile Sea Range	$ 471,579.94	3.2		Exercise Labor	$ 545,205.36
	Commercial Air Svcs	$ 663,305.30	6.1		Surge Support	$ 940,351.00
	Environmental	$ 73,081.00	5.1		Connectivity	$ 168,054.74
		$1,207,966.24				1,653,611.10
Joint/Coalition						
	USAF - AWACS	$ 343,719.00	3.2/5.1/6.1		USAF - AWACS	$ 34,600.00
					USAF - ACS/CRC	$ 3,775.00
					USAF - CAOC	$ 19,000.00
	USA - Patriot	$ 351,718.30	6.1		USA - Patriot	$ 96,722.00
	Germany	$ 1,587,600.00	6.2		Germany	$ 37,726.13
	UK	$ 1,460,592.00	6.2		UK	$ 31,153.00
		$ 3,743,629.30				$ 222,976.13

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

Live Totals	JMET		Synthetic Totals	JMET			
$ 540,323.74	3.2		$ 155,475.64	3.2		$ 25,573.12	$ 153,438.74
$ 244,940.50	5.1		$ 778,219.78	5.1			
$ -	5.2		$ 313,053.77	5.2			
$ 1,118,139.30	6.1		$ 552,768.91	6.1		127865.6167	
$ 3,048,192.00	6.2		$ 68,879.13	6.2			
$ 4,951,595.54			$ 1,868,397.23				
			$ 1,555,343.46	Total Excluding OP 5.2			
			$ 1,708,782.20				
Pct Breakouts	JMET		Pct Breakouts	JMET			
10.91	3.2		8.32	3.2			
4.95	5.1		16.76	5.1			
0	5.2		41.65	5.2			
22.58	6.1		29.59	6.1			
61.56	6.2		3.69	6.2			
100.00			100.00				

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

Direct Costs FST- J 07-1				
	JTEN Costs		$ 222,514.11	3.2/5.1/5.2/6.1/6.2
			$ 22,251.41	3.2
			$ 89,005.64	5.1
			$ 44,502.82	5.2
			$ 44,502.82	6.1
			$ 22,251.41	6.2

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

705th EXS (DMOC) - Exercise Planning/Tech Support	$2,400	$67,100	5.1
552nd OSS - Exercise Planners/DMOC E-3A Virtual Aircrew		$25,989	6.15
608th COS/CPS - USN CAOC Support		$13,731	6.1.5
31st ADA BDE - ADAFCO Team to CRC		$4,140	6.1.5
11th ADA BDE - Exercise Planning and ADAFCO Team to AEGIS Ship (Norfolk)	$7,600	$4,546	6.1.5
32nd AAMDC - Exercise Planning and LNO to CAOC	$2,855	$2,273	6.1.5
552nd OSS - AWACS LNO to CAOC		$2,668	5.2
107th ACS - USAF *Joint Simulation Cell*		$5,380	6.1.5
607th ACS - USAF Joint Simulation Cell		$13,250	6.1.5
109th ACS - Exercise Planning/DMOC CRC Virtual Aircrew	$7,395	$9,635	5.2
SH-60B MRT3		$31,500	6.2
EP-3E MAST		$31,000	5.1
E-2C EDRT		$23,875	6.1.5
NWDC On-Site Tech Support		$119,000	5.1
CSCS Tech Support (Mayport)		$4,200	5.1
Joint Air Defense SME (Mr. Stu Price)		$55,000	6.1.5
TOTALS	$20,250	$413,287	
		$413,287	

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

Personnel Costs FST -J 07-1	Description	Cost	JMET			
Exercise Labor	705th EXS (DMOC) - Exercise Planni	$2,400	3.2/5.1/6.1			
	11th ADA BDE - Exercise Planning and ADAFCO Team to AEGIS Ship (Norfolk)	$7,600	5.1/6.1.5		$1,589.25	3.2
	109th ACS - Exercise Planning/DMO	$7,395	3.2/5.1/5.2/6.1		$6,249.0	5.1
	32nd AAMDC - Exercise Planning and	$2,855	5.1/6.1.5		$1,848.75	5.2
	Joint Air Defense SME (Mr. Stu Price	$ 55,000.00	6.1.5		$65,563.00	6.1
				$75,250		
Surge Suport	NWDC On-Site Tech Support	$119,000	5.1		$19,009.0	3.2
	705th EXS (DMOC) - Exercise Planni	$67,100	3.2/5.1/6.1		$156,750.0	5.1
	107th ACS - USAF Joint Simulation C	$5,380	3.2/6.1.5		$0.0	5.2
	607th ACS - USAF Joint Simulation C	$13,250	3.2/6.1.5		$33,171.0	6.1
	CSCS Tech Support (Mayport)	$4,200	5.1			
				$208,930		
Joint/Coalition						
USAF - AWACS	552nd OSS - Exercise Planners/DMO	$25,989	3.2/5.1/6.1		$7,124.80	3.2
USAF - ACS/CRC	109th ACS - Exercise Planning/DMO	$9,635	3.2/5.1/6.1		$58,525.00	5.1
USAF - CAOC	552nd OSS - AWACS LNO to CAOC	$2,668	5.2		$2,668	5.2
USN - CAOC	608th COS/CPS - USN CAOC Support	$13,731	5.1/6.1.5		$49,539.20	6.1
USA - CRC	31st ADA BDE - ADAFCO Team to CR	$4,140	5.1/6.1.5		$31,500	6.2
USA - CAOC	32nd AAMDC - Exercise Planning and	$2,273	5.1/6.1.5			
					$433,537.00	
USN - ASW	SH-60B MRT3	$31,500	6.2			
USN - IO	EP-3E MAST	$31,000	5.1			
USN - AD	E-2C EDRT	$23,875	5.1/6.1.5			
USA - Patriot	11th ADA BDE - Exercise Planning and ADAFCO Team to AEGIS Ship (Norfolk)	$4,546	5.1/6.1	$149,357		
		$433,537		433537		

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

705th EXS (DMOC) - Exercise Planning/Tech Support	$2,400	$67,100	5.1				
552nd OSS - Exercise Planners/DMOC E-3A Virtual Aircrew		$25,989	6.15		Exercise Labor		
608th COS/CPS - USN CAOC Support		$13,731	6.1.5				$1,589.25
31st ADA BDE - ADAFCO Team to CRC		$4,140	6.1.5				$ 6,249.00
11th ADA BDE - Exercise Planning and ADAFCO Team to AEGIS Ship (Norfolk)	$7,600	$4,546	6.1.5				$ 1,848.75
32nd AAMDC - Exercise Planning and LNO to CAOC	$2,855	$2,273	6.1.5				$ 65,563.00
552nd OSS - AWACS LNO to CAOC		$2,668	5.2				
107th ACS - USAF Joint Simulation Cell		$5,380	6.1.5		Surge Support		
607th ACS - USAF Joint Simulation Cell		$13,250	6.1.5				$19,009.00
109th ACS - Exercise Planning/DMOC CRC Virtual Aircrew	$7,395	$9,635	5.2				$156,750
SH-60B MRT3		$31,500	6.2				$0
EP-3E MAST		$31,000	5.1				$33,171
E-2C EDRT		$23,875	6.1.5				
NWDC On-Site Tech Support		$119,000	5.1		Joint/Coalition		
CSCS Tech Support (Mayport)		$4,200	5.1				$7,124.80
Joint Air Defense SME (Mr. Stu Price)		$55,000	6.1.5				$58,525
TOTALS	$20,250	$413,287					$2,668
							$49,539
							$31,500.00

Appendix H – FY06 MCO Live and Synthetic Costs (continued)

			Overall Costs	
			$ 22,251.41	3.2
			$ 89,005.64	5.1
			$ 44,502.82	5.2
			$ 44,502.82	6.1
			$ 22,251.41	6.2
			$ 222,514.11	
			Total Cost	
			$ 656,051.11	
		FY Comparison		FST -J 06-1
$ 153,438.74	$ 25,573.12		OP 3.2	$ 181,048.76
$ -	$ -		OP 5.1	$ 778,219.78
$ -	$ -		OP 5.2	$ 313,053.77
$ -	$ 127,865.62		OP 1.2	$ 680,634.52
			OP 6.2	$ 68,879.13
				$ 2,021,835.97
		Overall ROI		
		JMET	3.08	
		OP 3.2	3.62	
		OP 5.1	2.51	
		OP 5.2	6.39	
		OP 6.1	3.53	
		OP 6.2	1.28	

Appendix I – Noble Resolve 07-2 Costs

					Synthetic						
Noble Resolve 07-2											
					$ 3,500,000.00						
	Exercise Labor				$ 2,450,000.00			Surge Suport			
		Scenario Development		$ 612,500.00					M&S Support		
Overall Apportionment forExercise Labor Support		OP 1.1	$ 30,625.00								OP 1.1
OP 1.1	0.05										
OP 4	0.1	OP 4	$ 18,375.00								OP 4
OP 5	0.55										
OP 7	0.3	OP 4.7	$ 42,875.00					Overall Apportion mentfor M&S Support			OP 4.7
	1							OP 1.1	0.05		
		OP 5.1	$ 168,437.50					OP 4	0.1		OP 5.1
								OP 5	0.55		
		OP 5.2	$ 67,375.00					OP 7	0.3		OP 5.2
									1		
		OP 5.5	$ 33,687.50								OP 5.5
		OP 5.7	$ 67,375.00								OP 5.7
		OP 7	$ 19,547.87								OP 7
		OP 7.3	$ 66,462.77								OP 7.3
		OP 7.4	$ 97,739.36								OP 7.4
			$ 612,500.00								
		Scenario Planner		$ 980,000.00							

Appendix I – Noble Resolve 07-2 Costs (continued)

OP 1						
1.1	Conduct Operational Movement					
OP 4	Provide Operational Logistics and Personnel Support					
4.7	Provide Politico-Military Support to Other Nations, Groups, and Government Agenecies					
OP 5						
5.1	Acquire and Communicate Operational Level Information and Maintain Status					
5.2	Assess Operational Situation					
5.5	Establish, Organize, and Operate a Joint Force Headquarters					
5.7	Coordinate and Integrate Joint/Multinationas and Intragency Support					
OP 7	Counter CBRNE Weapons in the Joint Operations Area					
7.3	Coordinate Passive CBRNE Defense in the Joint Operations Area					
7.4	Coordinate Consequence Management (CM) in Joint Operations Areas (JOAs)					
	Totals					
	OP 7.2	3		0.0638		0.1064
	OP 7.3	17		0.3617		
	OP 7.4	25		0.5319		
	OP 7.5	2		0.0426		
		47				

Appendix I – Noble Resolve 07-2 Costs (continued)

Notes:		1) Based upon the totals above, neither OP 7.2 nor 7.3 are significant for this exercise when viewing costs, even combined still only 10% of total. Costs will be assigned in ratio as shown above, with the .1064 being assigned to OP 7
		2) Information sharing was a major objective of all agencies involved and was a focus of the AAR comments and Lessons Learned, so will be apportioned to OP 5.1 as a larger portion of OP 5 assignments
		3) A stated ability to maintain a COP was voiced by 5/10 agencies, this will be factored into OP 5 breakout
		4) Coordination was a stated goal of 7/10 agencies involved which will be attributed to OP 4.7
		5) Other JMET addressed with the cooperation goal is OP 5.7
		6) The two overarching goals for NR 07-2 were stated as information sharing solutions applicable to CM and to combine real world and sim track data and display on a COP tool accessible Feceral, State, Local, Tribal, Intenational, Private and Non-Govermental
		7) Warfighter challenge issues map well to the following JMETs: #59 - OP 5.1and OP 5.7, # 123 - OP 5.5, OP 5.2, OP 4.7, and OP 7.4, #135 - OP 4, OP 4.7, OP 7, and OP 7.4,

Appendix I – Noble Resolve 07-2 Costs (continued)

Scenario Planner		$ 980,000.00
OP 1.1	$ 49,000.00	
OP 4	$ 29,400.00	
OP 4.7	$ 68,600.00	
OP 5.1	$ 269,500.00	
OP 5.2	$ 107,800.00	
OP 5.5	$ 53,900.00	
OP 5.7	$ 107,800.00	
OP 7	$ 31,276.60	
OP 7.3	$ 106,340.43	
OP 7.4	$ 156,382.98	
	$ 980,000.00	

Appendix I – Noble Resolve 07-2 Costs (continued)

Game Controller		$ 857,500.00
OP 1.1	$ 42,875.00	
OP 4	$ 25,725.00	
OP 4.7	$ 60,025.00	
OP 5.1	$ 235,812.50	
OP 5.2	$ 94,325.00	
OP 5.5	$ 47,162.50	
OP 5.7	$ 94,325.00	
OP 7	$ 27,367.02	
OP 7.3	$ 93,047.87	
OP 7.4	$ 136,835.11	
	$ 857,500.00	
	$2,450,000.00	$ 2,450,000.00

Appendix I – Noble Resolve 07-2 Costs (continued)

Total Resources by JMET			Percentage Expenditures by JMET
OP 1.1		$ 175,000.00	5.00
OP 4		$ 105,000.00	3.00
OP 4.7		$ 245,000.00	7.00
OP 5.1		$ 962,500.00	27.50
OP 5.2		$ 385,000.00	11.00
OP 5.5		$ 192,500.00	5.50
OP 5.7		$ 385,000.00	11.00
OP 7		$ 111,702.13	3.19
OP 7.3		$ 379,787.23	10.85
OP 7.4		$ 558,510.64	15.96
		$ 3,500,000.00	100.00

Appendix J – FA PANAMAX Trial Results

FA PANAMAX 07 Trials							
		Norm. Avg. (OE)	Fl Wt. Avg (OE)	Norm Avg(APE)	Fl. Wt. Avg. (APE)	LDF(Agg)	LDF (APE)
Day 1	Assess	0.720777	0.729638	0.753751	0.755604	0.530553	0.697962
	Plan	0.738289	0.741196	0.741228	0.742187	0.33631	0.606762
	Execute	0.737857	0.735213	0.777881	0.763292	0.427164	0.669216
Day 2	Assess	0.829977	0.824481	0.81875	0.827091	0.705142	0.773609
	Plan	0.824536	0.825865	0.831762	0.8321	0.429686	0.733846
	Execute	0.829949	0.822786	0.834631	0.837707	0.528257	0.755653
Day 3	Assess	0.857817	0.855466	0.866431	0.863156	0.768377	0.827932
	Plan	0.856028	0.852342	0.851641	0.85289	0.52668	0.762633
	Execute	0.860512	0.860179	0.854602	0.861632	0.583509	0.802272
Day 4	Assess	0.87218	0.872468	0.870941	0.861261	0.77789	0.824375
	Plan	0.870356	0.869024	0.858793	0.853383	0.583509	0.768364
	Execute	0.870356	0.869024	0.859907	0.85882	0.598296	0.782486

Appendix K – Observed Performance Results

Model Performance					
NR 07-2				LDF(APE)	LDF(LIM)
	OE	APE			
Assess	0.8177	0.7954		0.5869	0.8250
Plan	0.8158	0.8007		0.8380	0.7883
Execute	0.8125	0.7977		0.8491	0.7623
Assess	0.8465	0.8431		0.6218	0.8597
Plan	0.8405	0.8393		0.8631	0.8159
Execute	0.8375	0.8492		0.8645	0.8343

Appendix K – Observed Performance Results (continued)

TTGL CFMCC V3					
	OE	APE	LDF(APE)	LDF(LIM)	2nd Lvl
Assess	0.8029	0.7302	0.5647	0.7992	0.6092
Plan	0.8109	0.8228	0.7762	0.8048	0.5184
Execute	0.7747	0.7948	0.7120	0.7861	0.5054
Assess	0.8029	0.7560	0.6023	0.8219	0.6499
Plan	0.8109	0.8232	0.7762	0.8048	0.5792
Execute	0.7747	0.7911	0.7413	0.8115	0.5582
Assess	0.8029	0.7721	0.6231	0.8423	0.6863
Plan	0.8109	0.8222	0.7762	0.8048	0.6612
Execute	0.7747	0.7942	0.7671	0.8384	0.6346

Appendix K – Observed Performance Results (continued)

NR 07-2					
Day 2					
	OE	APE	APE(LDF)	Lim LDF	APE (LDF-2L)
Assess	0.8187	0.7901	0.5879	0.8265	0.6939
Plan	0.8247	0.8007	0.8380	0.7883	0.4026
Execute	0.8158	0.8285	0.8743	0.8389	0.6211
Assess	0.8187	0.7938	0.5675	0.8499	0.72908
Plan	0.8247	0.8119	0.8556	0.8162	0.46751
Execute	0.8158	0.8244	0.8844	0.8565	0.66503
Assess	0.8187	0.7898	0.5514	0.8873	0.7731
Plan	0.8247	0.8021	0.8749	0.8623	0.5722
Execute	0.8158	0.7842	0.8838	0.8665	0.6490

Appendix K – Observed Performance Results (continued)

NR 07-2					
Day 3					
	OE	APE	APE(LDF)	Lim LDF	APE (LDF-2L)
Assess	0.8440	0.8474	0.6242	0.8630	0.7723
Plan	0.8472	0.8393	0.8631	0.8159	0.4942
Execute	0.8424	0.8469	0.8785	0.85301	0.6132
Assess	0.8440	0.8466	0.5976	0.8813	0.79875
Plan	0.8472	0.8391	0.8782	0.8402	0.55235
Execute	0.8424	0.8457	0.8896	0.8700	0.6532
Assess	0.8440	0.8481	0.5753	0.9102	0.8319
Plan	0.8472	0.8402	0.8949	0.8803	0.6468
Execute	0.8424	0.8471	0.8953	0.8846	0.6789

Appendix K – Observed Performance Results (continued)

Initial Bins					
	OE	APE	LDF(APE)	LDF(LIM)	2nd Lvl
Assess	0.8332	0.8289	0.5998	0.7212	
Plan	0.8305	0.8283	0.5660	0.7144	
Execute	0.8336	0.8244	0.5629	0.7168	
Assess	0.8332	0.8283	0.7059	0.7833	
Plan	0.8305	0.8259	0.6722	0.7632	
Execute	0.8336	0.8239	0.6937	0.8246	
Assess	0.8332	0.8284	0.7362	0.8097	
Plan	0.8305	0.8245	0.7060	0.7929	
Execute	0.8336	0.8239	0.7220	0.8456	

Appendix K – Observed Performance Results (continued)

RDD NR 07-2				
	OE	APE	APE(LDF)	Lim LDF
Assess	0.8440	0.8474	0.6242	0.8630
Plan	0.8472	0.8393	0.8631	0.8159
Execute	0.8424	0.8469	0.8785	0.85301
Assess	0.8440	0.8466	0.5976	0.8813
Plan	0.8472	0.8391	0.8782	0.8402
Execute	0.8424	0.8457	0.8896	0.8700
Assess	0.8440	0.8481	0.5753	0.9102
Plan	0.8472	0.8402	0.8949	0.8803
Execute	0.8424	0.8471	0.8953	0.8846

Appendix K – Observed Performance Results (continued)

RDD TOPOFF 04				
	OE	APE	APE(LDF)	Lim LDF
Assess	0.9050	0.9500	0.9790	0.9790
Plan	0.9016	0.9330	0.8427	0.8427
Execute	0.9170	0.8986	0.9271	0.9430
Assess	0.9050	0.9500	0.9781	0.9781
Plan	0.9016	0.9310	0.8427	0.8427
Execute	0.9170	0.8975	0.9358	0.9504
Assess	0.9050	0.9500	0.9791	0.9791
Plan	0.9016	0.9278	0.8427	0.8427
Execute	0.9170	0.8900	0.9388	0.9388

9 VITA

MARK CULLEN NESSELRODE

DEGREES:

Doctor of Philosophy (Engineering with a Concentration in Modeling and Simulation), Old Dominion University, Norfolk, VA, May 2008

Degree of Mechanical Engineering, Naval Postgraduate School, Monterey, CA, March 1985

Master of Science in Mechanical Engineering, Naval Postgraduate School, Monterey, CA, March, 1985

Bachelor of Science (Mathematics), United States Naval Academy, Annapolis, MD, June 1976

PROFESSIONAL CHRONOLOGY

Virginia Modeling, Analysis and Simulation Center, Old Dominion University, Norfolk, VA, Research Intern, August 2006 – Present

Tactical Training Group Atlantic, Dam Neck, VA, Commanding Officer, September 2003 – June 2006

USS ANZIO (CG-68), Norfolk, VA, Commanding Officer, August 2001 – August 2003

USS OLDENDORF (DD-972), San Diego, CA, Commanding Officer, February 1995 – October 1996

SCHOLARLY ACTIVITIES COMPLETED:

Nesselrode, M.C. Developing a Repeatable and Reliable Methodology for Determining Return on Investment. In Proceedings of the first Conference on Mastering Engineering Complex Situations (MECS) 22-24 October 2007, Norfolk, VA

Nesselrode, M.C. Developing a Repeatable and Reliable Methodology for Determining Return on Investment in s Synthetic Environment, In Proceedings of the 75th Military Operations Research Society Symposium (MORSS), June 12-14, 2007, Annapolis, MD