Report of the
Defense Science Board
Task Force
on

Improving
Test and Evaluation
Effectiveness

December 1989

Office of the Under Secretary of Defense for Acquisition
Washington, D.C. 20301-3140

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp. Date Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution Statement A. Approved for Public Release. Distribution is Unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) N/A	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a. NAME OF PERFORMING ORGANIZATION Defense Science Board, Ofc of the Under Secy of Def (A&T) 6b. OFFICE SYMBOL *(If applicable)* DSB/OUSD (A&T)	7a. NAME OF MONITORING ORGANIZATION N/A
6c. ADDRESS *(City, State, and ZIP Code)* The Pentagon, Room 3D865 Washington, DC 20301-3140	7b. ADDRESS *(City, State, and ZIP Code)* N/A
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Defense Science Board, OUSD (A&T) 8b. OFFICE SYMBOL *(If applicable)* DSB/OUSD (A&T)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N/A
8c. ADDRESS *(City, State, and ZIP Code)* The Pentagon, Room 3D865 Washington, DC 20301-3140	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: N/A PROJECT NO.: N/A TASK NO.: N/A WORK UNIT ACCESSION NO: N/A

11. TITLE *(Include Security Classification)* Report of the Defense Science Board Task Force on Improving Test and Evaluation Effectiveness, December 1989, Unclassified.

12. PERSONAL AUTHOR(S)
N/A

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Final	FROM N/A TO N/A	1989 December	73

16. SUPPLEMENTARY NOTATION
N/A

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	

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

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Diane L.H. Evans	22b. TELEPHONE *(Include Area Code)* (703) 695-4157/8 22c. OFFICE SYMBOL DSB/OUSD (A&T)

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

This Document Has Been

CLEARED

For Open Publication

Directorate for Freedom of Information
and Security Review, OASD(PA)
Department of Defense

BLANK PAGE

OFFICE OF THE SECRETARY OF DEFENSE
WASHINGTON, D.C. 20301-3140

DEFENSE SCIENCE BOARD

2 8 DEC 1993

MEMORANDUM FOR THE SECRETARY OF DEFENSE
THE UNDER SECRETARY OF DEFENSE

SUBJECT: Final report of the Defense Science Board Task Force on Improving Test and Evaluation Effectiveness

Attached is the final report of the Defense Science Board Task Force on The Contributions of Modeling and Simulation (M/S) to Defense Test and Evaluation, chaired by BGEN Robert A. Duffy, USAF (Ret.). This report highlights a number of significant steps regarding the use of models and simulations that can result in current and future improvements in test and evaluation.

The Task Force determined that M/S can help provide more illumination of choices in the operational requirement process, increase flexibility in the development process and assist in the preparation of early operational assessments. As an example, by placing more emphasis on early and continual operational evaluation during development, operational problems can be identified early.

A general task force consensus is that a process needs to be established that translates the operational requirements into an evaluation framework. Models and simulations are expected to play a key role in the development of this framework. At present, the translation of requirements to technical criteria and then into an evaluation framework is judged to be ambiguous.

I suggest that you read the attached letter from the Chairman, the Executive Summary and recommendations, and approve the report for publication.

Robert R. Everett
Chairman

Attachments
as

OFFICE OF THE SECRETARY OF DEFENSE
WASHINGTON, D.C. 20301-3140

DEFENSE SCIENCE BOARD

14 DEC 1989

Mr. Robert R. Everett
Chairman
Defense Science Board

Dear Mr. Everett:

Attached is the final report of the Defense Science Board Task Force on Improving Test and Evaluation Effectiveness. The Task Force identified a number of potential uses of models and simulations to improve test and evaluation and the acquisition process if employed early and effectively in the system life cycle. The use of models and simulations can amplify and expand our understanding of system and mission requirements, system effectiveness, and costs resulting from acquisition decisions.

To achieve this potential a need exists for a process that will provide an independent, objective evaluation of model and simulation utilization. Credibility of models and subsequent simulations are important and must be considered in light of their application to a specific problem.

The Task Force has recommended several significant and broad actions to improve test and evaluation and the acquisition process:

o Support early involvement of the operational test community in the development phase through a process that uses an evaluation framework.

o Use simulations to help determine the events and criteria that must be tested.

o Conduct excursion and sensitivity analyses to focus on system engineering criteria that validates modeling results and contributes to an early operational assessment.

o Support acceptance of simulation as an evaluation tool by increasing development phase flexibility through a process that allows re-evaluation as threat, technology and knowledge evolve.

o Involve users early with mock-ups of man/machine interfaces to enable a better understanding of the system design.

I want to thank all of the members of this panel for their contributions to this report.

Sincerely,

BGEN Robert A. Duffy, USAF (Ret.)
Chairman, DSB 1989 Summer Study Task
Force on Improving Test and
Evaluation Effectiveness

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Over the last several years, models and simulations have been increasingly used to support the development, test and evaluation process for new weapon systems. This practice is expected to continue. Over the last ten years we have seen dynamic growth in the computing and networking technology areas, which are the underpinning for digital simulation. This trend is expected to continue, and will permit lower cost and higher fidelity simulations. As a result, the Task Force on Improving Test and Evaluation Effectiveness was requested to look at ways of improving the use of modeling and simulation as tools in the test and acquisition of defense systems.

A number of significant steps were identified regarding the use of models and simulations that can result in immediate and future improvements in test and evaluation. Modeling and simulation can be an effective tool in the acquisition process throughout the systems life cycle, but most importantly if employed at the inception of the system's existence. Modeling and simulation can contribute knowledge and understanding of system and mission requirements, system effectiveness, and costs resulting from acquisition decisions, but only if used properly.

The task force came to several conclusions with regard to the use and credibility of models and simulations in the test and evaluation and acquisition processes:

- o The foundation of the acquisition process, the operational requirement and its translation into system terms, can be improved through the use of modeling and simulation.
- o An early and continual involvement of the OT&E community in the requirements process can contribute and improve the acquisition process.
- o A development program, as embodied in a specification and contract, may become overly rigid, restricting the willingness to evaluate and incorporate changes as threat, technology and knowledge evolve. Modeling and simulation can be used as a tool for continual evaluation of potential changes.
- o In cases where modeling and simulation raises items of uncertainty in terms of system requirements to achieve operational utility, unanticipated early operational tests may be warranted even while a system is still under development.
- o Accounting for human performance early in system acquisition improves system capability and enhances the ability of the test and evaluation process to predict operational performance.
- o The availability of high quality, reusable models and simulations could decrease redundant efforts while improving quality of key elements.

The credibility of a model cannot be considered separately from its application to a specific problem, validity of data inputs, and qualifications of those executing the model and interpreting the results. Current DAB documentation does not address model and simulation credibility. Advanced technology in both hardware and software offer opportunities for improving the credibility and applicability of models and simulations, but continued research is needed. In view of the limitations of models and simulations, the approach for effective use of modeling and simulation in the operational requirements area should be one of identifying areas where uncertainty levels require early operational testing. Since operational testing is expensive, the isolation of areas which require tests is important.

Early in the course of this study, the task force discovered that, in order to make useful suggestions on modeling and simulation, corresponding recommendations in the areas of test and acquisition would likewise be necessary. The task force has recommended several significant and broad suggestions to improve test and evaluation and the acquisition process:

- Emphasize the cooperative learning role for operational testing during development, and support this activity through a process that uses an evaluation framework established at the start of a program.

- Do not employ simulations to prove or disprove things, but instead exploit their ability to isolate high sensitivity areas. Simulation has an important role in providing sensitivity analyses, and as a method of focusing on system engineering issues early through operational tests.

- Confidence in models can be enhanced by employing them for excursion and sensitivity analyses, and focusing on critical issues by running tests and validating the results. It is not feasible or cost effective to set up a central office to accredit modes, nor is it necessary to implement a management process to distribute and reuse simulations.

- OSD must allow the development phase to become less rigid and support the acceptance of simulation as an evaluation tool.

- The tools are available and the cost is sufficiently low such that every program should build mock-ups of man/machine interfaces as early as possible, and involve actual users to better understand the utility of the system design.

This study has determined that it is important to anticipate operational test issues, both for reasons of cost and credibility. More emphasis on operational testing is needed during system development so that operational problem areas can be identified while they are still economically resolvable. A process must be established that defines evaluation frameworks which predict probable evaluation procedures; then, as the real program progresses, the frameworks should be upgraded consistent with the advancing state of knowledge. Also, it must be acknowledged that the current acquisition process stifles evaluation, and it is recommended that upper management levels provide direction to develop more open attitudes regarding the responsibilities and contributions of the test and evaluation communities in the acquisition process.

Finally, the task force found no need to establish an independent agency or office to accredit or manage the use or distribution of simulations. Modeling and simulation can and should be used to focus testing into those functional and operational areas where there is a lack of assurance, and to recognize those areas where sensitivity is sufficiently questionable that actual testing is in order. In this way confidence in the final product is realized - through testing married with simulation.

SECTION 1 - INTRODUCTION

Recent demands to reduce DoD spending and costs associated with systems acquisition for both hardware and software development activities have prompted a look at the acquisition life cycle. More specifically, the phases of development and operational testing continue to reflect high visibility as critical points to assess system credibility. The possibility that the addition of modeling and simulation to the process might provide valuable insights was raised. As a result, the Defense Science Board was tasked (Terms of Reference, Appendix A) to study how to improve the effectiveness of test and evaluation with modeling and simulation (M/S) as the focus. The tasking specifically requested the DSB to:

- o Review prevailing use of models, laboratory tests and field tests.
- o Determine appropriate situations in which to test and/or model.
- o Determine the required fidelity of representation of the system under test and its environment.
- o Determine which discipline will govern the interpretation of results.

The Task Force membership, Appendix B, heard a variety of presentations, which are listed in Appendix C.

MODELS AND SIMULATIONS

Models and simulations have been and continue to be used extensively to support the weapon development, test and evaluation process. Such use can be expected to continue - if not increase - in the future.

Defense systems are increasingly complex. Their operational utility depends increasingly on successful performance at extended ranges. The integration of sensor information from multiple parts of the electromagnetic spectrum or from multiple sources is becoming a significant factor. The systems are required to operate in adverse environments (weather, hostile electromagnetic, space, enemy countermeasures). To be effective, they must interact with other systems, often over great distances. Advanced command and control systems are required to overcome these difficulties. Furthermore, systems now being developed are expected to remain effective in the future against threats that will evolve in ways not fully predictable. A complete test of such systems would be large in scope and require the generation of conditions that are difficult if not impossible to create short of actual combat. The practicalities of cost, test range space, availability of advanced threat systems/surrogates, safety, etc., will necessarily limit test data availability. Forming an overall evaluation of a major system's performance will almost always require a "model" - even if only a mental one - to integrate the available test data and to extrapolate to those conditions which cannot be created in the test environment. These models and simulations are not replacements for test data, but rather complementary tools in the evaluation process.

In the broadest sense, a model is a representation of an object, system, or process (or a part thereof) in mathematical, physical or logical terms, usually simplified, often idealized or abstract, serving as a basis for calculations, predictions or further investigations. Simulation is a technique for experimentation in which the operation and dynamics of a real-world system are imitated or reproduced by some different system, usually involving one or more models.[1] Under these definitions, a scaled representation of an aircraft is a model; placing that representation in a wind tunnel to study flight dynamics would be a simulation.

The basic concept underlying these definitions is that a model is some abstraction that embodies our understanding of a system while a simulation represents the dynamic exercise of one or a set of models.

A large variety of simulations have been developed to support different aspects of the development and evaluation process. Simulations differ as to the scope of the "system" they attempt to reproduce. Simulations are available at the system/subsystem level which typically are based on detailed models of the relevant physics or engineering of the system. Combat simulations can be at the one-on-one level, few-on-few, many-on-many, battalion, theater, etc. Generally, these simulations sacrifice detail in the modeling of individual systems in return for an increase in scope.

A simulation can be only as good as the knowledge incorporated in the models it exercises. Since the ability to model human performance and decisionmaking is very limited, the degree to which modeling of human factors is needed and the manner in which this is attempted is an important characteristic of a military simulation. Physical and engineering models may not require it. Many simulations are analytic in that human factors are accounted for by some set of algorithms or decision rules which stay fixed during the simulation run. Interactive simulations exist which allow differing degrees of human involvement. Some require higher level decisions to be made by humans (e.g., tactics of a unit), while the modeling at the individual platform/weapon level does not. Manned simulators replicate systems in greater detail and require a person to "operate" the individual system. Simulators can incorporate the actual operating software of subsystems. Manned simulators for both air and ground systems are being increasingly internetted. Hardware-in-the-loop simulation is an analogous technique wherein the actual hardware is made to operate by a simulated stimulus.

The variety of simulation types and the number of individual simulations which have been developed reflect, in part, the range of evaluation decisions they are expected to support.

[1] These "definitions" have been synthesized from multiple sources including technical and general purpose dictionaries and the definitions given in DoD Directive 5000.3.

SECTION 2 - FINDINGS

The Acquisition Process

Effective systems development involves processes like planning, analysis, decision coordination, requirements definition, funding, design, development, fabrication and test and evaluation. This set of activities we term the "acquisition process".

The acquisition process can function in a number of ways depending on the focus of the organization tasked with the responsibility to develop the military system.

Early in the acquisition process many of the activities focus on the development of concepts, requirements, and preliminary designs that are not well defined or understood. There are uncertainties and ambiguities that arise as the system concept is formulated with the focus toward operational utility. This part of the process tends to be "loose" in the context of an open set of sciences and technologies that are addressing the problem of defining operational requirements.

As illustrated in Figure 2-1, the notion of the process being "loose" at the front-end of the system development is characterized by: 1) the lack of experience with formal methods to define accurately system requirements; 2) a communication gap between the operational and engineering communities in the translation of complex requirements, 3) insufficient availability of analysis tools to assist in the formulation of requirements; and 4) the absence of the OT&E community as an observer to the front-end activities of the program, which tends to cause surprises when viewed at the end. Transition through the development phases causes the ambiguous capabilities, that were derived as requirements, to become rigid and fixed in the form of technical specifications and contract language which has been dictated by precedent and regulation.

Programs typically remain many years in the development process. The longer a program remains in the development cycle the more likely that changes to mission requirements will occur requiring reevaluation of the fundamental concepts and tradeoffs that underlie the technical specifications. The threat that generated the need in the first instance is a moving target.

The current process leaves little flexibility (i.e., as "rigid" baselined specifications) for integrating changes to systems under development, when those changes are caused by new technology or national priority or threat-motivated changes to mission requirements. While baselining is a positive step to produce program stability for long lasting programs, it can stifle needed changes. When changes are applied they add cost to contracts, adversely affect schedules, and can cause contractor or government deviations from their contractual commitments.

Recent changes to DoD standards and policies have streamlined the acquisition reporting chain and may simplify the generation of technical development requirements. Even with the implementation of current changes to the acquisition process, however, the potential for delay remains where changes must be introduced. Unless development timelines can be reduced or flexibility can be provided in the acquisition process, the potential for missed technology and delayed program changes will increase acquisition costs and reduce our ability to respond to the evolving threat.

As the system transitions to the operational community for testing, the realization that the operational utility of the system may be deficient comes too late for changes to be rapidly and economically applied to correct the system.

IMPROVING T&E EFFECTIVENESS

Acquisition Process

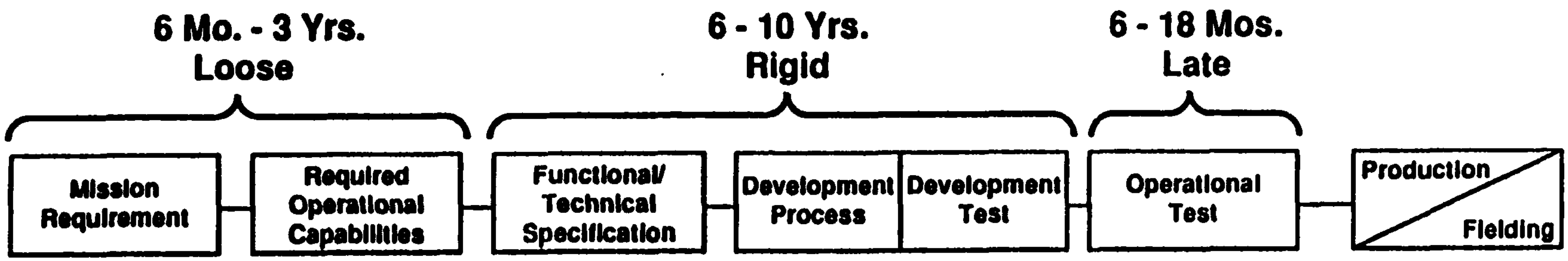


Figure 2-1

IMPROVING T&E EFFECTIVENESS

Role of T&E in the Acquisition Process

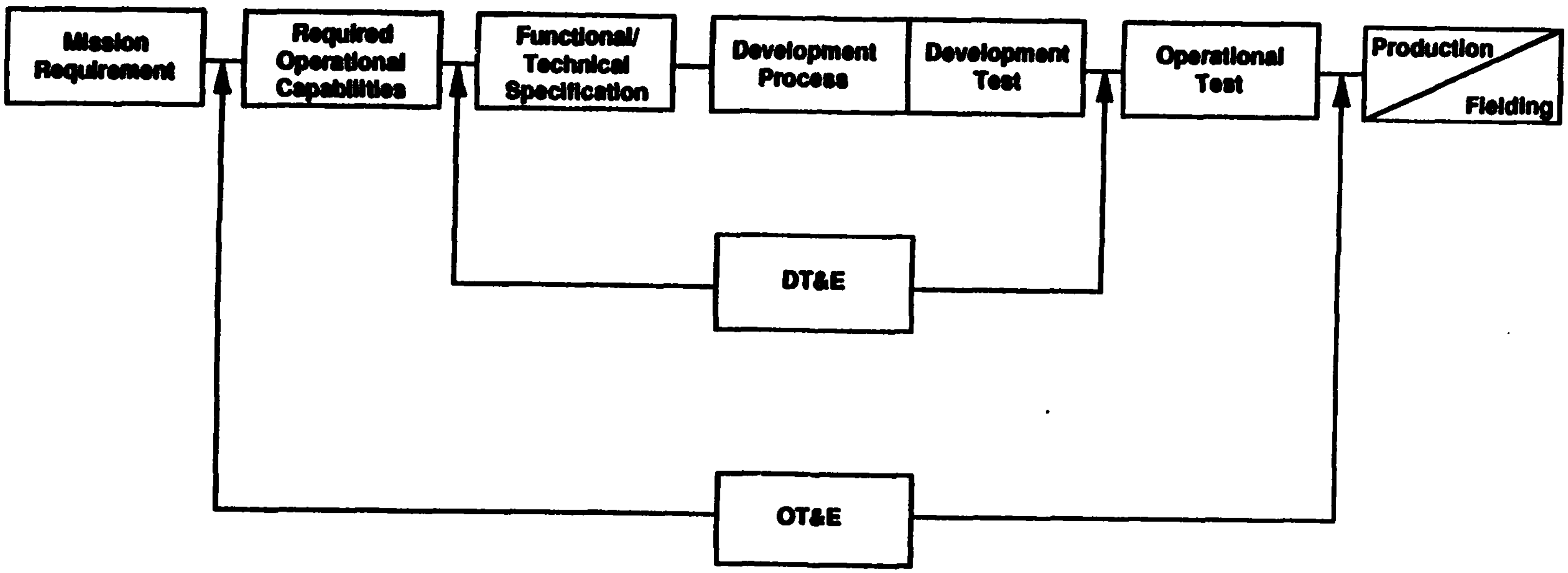


Figure 2-2

Operational testing returns focus to the mission. The current state of the system requirements and the ability of the developed system to attain operational utility against that requirement is audited. With regard to the original mission requirements, the result of system changes in need, understanding, and people in responsible positions, create new measures of effectiveness at a very late stage. This activity frequently creates surprises and may well require costly changes in the system during or after deployment.

The Role of Test and Evaluation in the Acquisition Process

As stated in DoD Directive 5000.3, "The primary purpose of all T&E is to make a direct contribution to the timely development, production, and fielding of systems that meet the users' requirements and are operationally effective and suitable." It is generally agreed that this should be accomplished by a continuous assessment of the system's capabilities (or potential capabilities) as it progresses through the process. Defense Test and Evaluation is organized into two parts: Development T&E and Operational T&E. DT&E is conducted to assist in the engineering design of the systems as well as to verify attainment of technical performance specifications, objectives and supportability (as identified in the contract between the Government and the Contractor). OT&E is conducted to determine the operational effectiveness and suitability of the system for use in combat by typical military users.

Test and evaluation is a critical component of our existing acquisition process. The test and evaluation world is split into two general communities of development and operational test and evaluation. Each plays a distinct role in the acquisition process, as pictured in Figure 2-2, but both exist to help field operational weapon systems that work and are effective in combat conditions. DT&E is focussed on making the system work and OT&E is focussed on how well it works.

DT&E

DoD 5000.3 states that DT&E is conducted throughout various phases of the acquisition process to ensure the acquisition and fielding of an effective and supportable system by assisting in the engineering design and development and verifying attainment of technical performance specifications, objectives, and supportability. DT&E is an integral part of the full-scale development process. They are constantly reviewing the design, prototypes, and development test results against the functional and technical specifications.

Development testing covers a wide range of components and conditions ranging from material sample testing to full up system testing. The purpose of development testing includes tests to evaluate design approaches, tests to collect data to validate analytical models, tests to assess technical risk, tests to verify technical performance, tests to demonstrate specification compliance, and tests to predict operational performance. All these tests are focussed on ensuring that the development process yields a system that complies with the technical specifications.

Development evaluation has historically focussed on evaluating the results of development testing against the requirements outlined in the technical specifications. Evaluation planning most often follows the test planning effort, thus the evaluation methodology is most often driven by the test events already planned. The evaluator must make do with the results available from the development tests.

OT&E

Title 10 U.S.C. Section 138 define OT&E as: "the field testing, under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for the purpose of determining the effectiveness and suitability of the weapons, equipment, or munitions for use in combat by typical military users; and the evaluation of the results of such tests."

OT&E is tasked with field testing weapon systems in realistic conditions to determine effectiveness and suitability. OT&E is the key player in the operational testing portion of the acquisition process. They perform the final examination of the weapons system to determines its effectiveness against the Required Operational Capabilities (ROC).

As prescribed by law, operational testing must test as much of the weapon system as possible in conditions as near as are achievable to actual combat. This mandate is difficult to achieve because of constraints due to cost, security, safety, test instrumentation, terrain, treaty compliance, and many others. For example, one cannot kill soldiers to determine the kill effectiveness of a new bullet or rifle design. The law also requires operational testing to determine operational effectiveness and suitability. Operational effectiveness as defined in DoD 5000.3-M-1 means: "The overall degree of mission accomplishment of a system when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, tactics, survivability, vulnerability, and threat (including countermeasures and nuclear threats)."

Operational suitability is defined as: "The degree to which a system can be satisfactorily placed in field use, with consideration given to availability, compatibility, transportability, reliability, wartime use rates, maintainability, safety, human factors, manpower supportability, logistic supportability, documentation, and training requirements."

The determination of operational effectiveness and suitability as defined above is impossible based solely on field test results without further analysis. This further analysis or evaluation must rely more and more on modeling and simulation as weapon systems increase in complexity and in light of the various constraints placed on field testing.

TEMP

The Test Evaluation and Master Plan (TEMP) defines and integrates the DT&E and OT&E efforts for all major weapon system procurements. It relates program schedule, decision milestones, test management structure, and test resources to critical technical characteristics, critical operational issues, evaluation criteria and procedures. It is used as a tool for oversight, review, and approval of the test and evaluation effort by OSD and all DoD components. The TEMP as described in DoD 5000.3-M-1 is brief by directive and explicitly covers the system requirements, program summary, DT&E, OT&E, and test and evaluation resources. The initial TEMP must be submitted to ODDDRE(T&E) prior to Milestone I and updated at least annually thereafter. In summary, the TEMP is viewed as a living document throughout the acquisition cycle, outlining the roles of DT&E and OT&E.

Two major weaknesses were observed in reviewing the role of test and evaluation in the acquisition process. The first observation was that there is a need to have the OT&E community participate throughout the entire acquisition process, not just at the end (as related to the "late" discussion earlier). The second observation was that the test and evaluation efforts rely almost solely on test results which often do not arrive until late in the development cycle when design changes are costly (as related to the "rigid" discussion earlier). These observations can be addressed and their effect mitigated by policy changes and increased use of modeling and simulation.

OT&E Participation

It was observed that the OT&E community is not heavily involved in the acquisition process until near the end of the program. As described earlier, the Task Force felt that this is too late to be of benefit. A weapon system tends to meet its technical specifications yet at times fails its operational tests. This is a fundamental weakness in the acquisition process and is attributed to unforecasted and perhaps unforecastable changes in the threat or environment. Early and continual involvement of the OT&E community (or its function) could help mitigate the effect of these surprises, reduce weapon system costs, and yield more effective weapon systems.

Test Emphasis

It was observed that the test and evaluation community places a heavy emphasis on test and a light emphasis on evaluation. Test and evaluation are interrelated and complementary processes, both of which are necessary; neither alone is sufficient. Evaluation can be used to judge overall system performance against the operational mission requirements and to reassess performance as the mission requirements and system design evolve. This evaluation is supplemented by test results. Evolving analytical models and simulations can help. A consistent and traceable set of evaluation tools could be used throughout the acquisition cycle to help mitigate surprises encountered during operational tests. The framework for the test and evaluation process could be documented and updated in the TEMP. Modeling and simulation could play a major role in improving the evaluation process.

The Role of M/S in the Acquisition Process

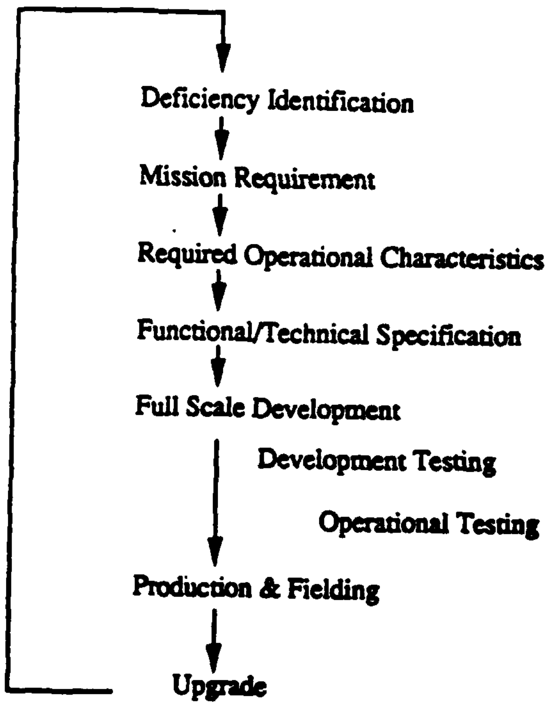
Modeling and simulation (M/S) is used extensively throughout the acquisition process. M/S is at times used in establishing the mission requirements, in designing the weapons system, and in forecasting the weapon system costs.

Use and Users of Simulations

Within the defense establishment, the body of simulation users within the Services and among defense contractors is large. These include Defense agencies, Service laboratories and schools, research centers, analysis divisions, program offices (PMs/SPOs), contractors and, more recently, the Service Operational Test Agencies (OTAs). Indeed, the list of those organizations not using simulations is probably quite short.[2]

[2]If simulations to support exercises, training and wargaming were considered, most of the field commands would be included, making the list of non-users vanishingly small.

Despite this myriad of users, simulation types/applications and the differences between Service organization/procedures, some generally applicable observations can be drawn of the acquisition process:



The process of identifying deficiencies and developing mission requirements is, of necessity, supported by limited and imprecise data. It is based largely on "theoretical" data which includes estimates of technological advances and their military applicability as well as estimates of threat advances and potential. "Historical" data - past combat results and performance of current systems - are also available. While physical/engineering simulations play an important role, translation of these data into a mission level requirement demands that some higher level "model" be used. If simulations are used, they have been operational (or higher) level simulations.[3] Typically they have also been analytic in nature. The same models also typically support the tradeoff analysis in COEAs[4] which support the Milestone[5] I decision at which time a "preferred" system approach is selected and its required operational characteristics are first defined. Particularly if Milestones 0 and I are not combined, prototype or demonstration/lab data

[3]In some acquisition programs, documented use of simulation to identify requirements is absent.

[4]Cost and Operational Effectiveness Analyses.

[5]Defense Acquisition Board (DAB) program decision point approving to proceed with next life-cycle phase, DoD Instruction 5000.2.

may also be available. The mission area analyses and COEAs that support the process of requirements are the responsibility of the Service staffs - i.e., the proponents of the system.

As the program moves through concept demonstration/validation and the program office becomes established, simulation tends to drop down to the system or engagement level (i.e., one-on-one). As technical performance data become available, engineering models are refined and then used as a development tool. These system or engagement level models are often used to quantify the government's Request for Proposals (RFP). Contractor proposals use system, engineering and cost models. A set of technical specifications for full scale development results. After Milestone I, simulation use resides increasingly in the program office and more significantly the contractor(s).

After the technical specifications have been placed in a contract and FSD progresses, the majority of simulation supports the attainment of those specifications rather than addressing the mission requirement directly. Uncertainty as to how well the mission requirement is being met can develop if the FSD system is not fully meeting the technical specifications. System and engineering level simulation dominates. Manned simulators are used extensively, especially in the aircraft industry. Hardware-in-the-loop simulations are also used, particularly in sensor/ECM programs. These simulators may also be used in supporting development testing. Engagement level modeling is not that uncommon, particularly if an engagement level requirement is in the contract (e.g., MLRS/TGW requirement of defeating X percent of a battalion with a salvo).

As the program progresses to the operational test phase, simulations continue to be used. Since operational tests are conducted in the field by troops, live munitions cannot be used. Therefore, simulators replicate firings (e.g., lasers and laser detectors) and simulations provide "kill" results. These simulations may be very detailed engineering models, such as missile fly-out models or P_k tables, which have been generated by vulnerability simulations off-line to the test. The fielded threat in operational tests also consists of "simulators", either U.S. systems of varying fidelity to those of the threat or specially designed manned simulators used as threat surrogates. Thus, a substantial amount of field data is based on simulation.

The Service operational test agencies (OTAs) that have the responsibility of conducting tests have traditionally not developed or used simulations in their evaluations. In the past two years, however, this has been changing as the Director of Operational Test and Evaluation has been urging that operational assessments be made early, well before the test itself. Since this is a relatively new effort and the OTAs have limited resources, prominent examples of simulation use in support of early operational assessments are lacking.

While the above simplified characterization applies to the "typical" system, some systems, at OSD direction, are also tested and evaluated in joint test and operational utility evaluations (OUEs). Some of these activities use simulation extensively. For example, the AMRAAM OUE was conducted on netted manned simulators.

Credibility of Models and Simulations

As the use of simulations has increased in DoD, the simulations themselves have grown in size and complexity. In fact, most simulations that appear in the acquisition process are too complicated to be sufficiently understandable by decisionmakers in the time available, leading to serious concerns of validity. It is becoming ever more important to ensure that the results of simulations used to support major acquisition decisions are reliable, and that they do not pose inordinate risks. Since simulations may be the only practical evaluation tools available for certain acquisition decisions, steps must be taken to assure a justifiable measure of confidence in the results provided by such models and simulations.

An essential attribute of every truly useful model or simulation is that it has earned a high degree of credibility; its construction, execution, and the interpretation of results are considered to be "good and true", taken in the proper context. In several previous studies of modeling and simulation, the credibility of a simulation has been identified as the key determinant of utility.

System and subsystem models at the engineering level, particularly those which model functions that are exercised in peacetime, seem to enjoy a fairly high level of credibility. For example, it is virtually impossible to imagine a modern aircraft development program that would not make extensive use of wind tunnels and flight dynamics simulations. It has been shown that these simulations frequently permit a reduction in the number of flight hours required during the development process. This high level of simulation credibility can be attributed to the degree of understanding of aerodynamics (at least empirically) and the use of instrumented flight test data to continually improve the fidelity of such simulations.

Complex combat simulations which estimate operational performance at the force-on-force level (some would also argue at the one-one-one engagement level) naturally encounter a great deal more skepticism, since these high level models must necessarily make simplifying assumptions and sacrifice detail. Contributing to this distrust is the fact that the fundamental theoretical bases for the simplifications are less well understood (Lanchester's equations hardly inspire the confidence of Maxwell's). The importance of human performance factors is an additional complication. It could be argued that only actual combat, with instrumentation to collect data, could fully resolve all the suspect elements of simulation. Use of multiple models with different theoretical approaches and assumptions may provide a hedge against the uncertainty of our fundamental knowledge of combat processes. These different approaches and assumptions must be made clear, however, else the different models are likely to generate more confusion than insight.

Although many attempts have been made to develop procedures for assessing the credibility of a model/simulation, none have gained widespread acceptance. At the present time, there is no policy or process in place in DoD to assess the credibility of specific models and simulations to be used in the test and evaluation and the acquisition processes.

In general, the Task Force notes that, due to the variability of applications and different issues to be addressed by simulation runs of similar models, it is unrealistic to attempt to "accredit" any single model for more than one specific application scenario. This is not to say, however, that a more rigorous evaluation process of model simulations is not needed; it is. Further, because of the wide diversity of models and simulation scenarios, it is unreasonable to expect that any single "accreditation" agency could be effective in evaluating every model proposed by DoD agencies in support of arguments related to acquisition decisions. Indeed, the Task Force holds that the credibility of a model or simulation has meaning only in the context of the model's application to a specific problem, and reflects not only the integrity of the model formulation itself, but also the validity of the input data and the qualifications of those executing the model and those interpreting the results. In short, the entire process that uses the results of simulations in arriving at conclusions should be subjected to scrutiny; the approval of any subset of procedural steps is insufficient to assure the credibility of results.

Most of the large simulation models used in DoD contain one or more of several "building block" submodels or databases developed independently by executive agents, such as the nuclear effects submodels formulated by DNA and threat submodels constructed by DIA. The credibility of these special purpose, reusable simulation modules derives primarily from confidence in the originating agency, and measures should be taken to support the currency, maintenance, and appropriate use of these submodules.

In contrast, the issue of credibility must be addressed anew for each unexplored application based on the larger, more complicated simulation models involving interactions beyond well-understood physical laws. Many of the more complex models have earned respect over time; however, caution is appropriate even for these cases since such models are used by disparate groups who may not be intimately familiar with the original constructs and assumptions of such models. The Task Force has found no standard process for ensuring credibility. As warranted by the import of the decision involved, close scrutiny of the modeling/simulation process and its application to the question at hand is needed, and may be best carried out by an independent panel of experts, selected on a case-by-case basis.

Although model simulations are frequently cited in support of proposed major acquisition programs, there is currently no prescribed reporting requirement that adequately addresses the use of modeling and simulation in the acquisition program, either prior to or during the development phases. The Task Force believes that this simple omission detracts from the credibility of properly employed simulations. The planned use of modeling and simulation should be reported in the Milestone documentation provided the DAB, at least in summary form, along with the methodologies of application and interpretation, a description of applicable simulation limitations, assumptions, extrapolations, and sensitivities.

Recent advances in technology have accelerated and broadened the application of modeling and simulation. Developments in the computer sciences can do much to promote the standardization of model simulation software structures, module and database interfaces, and languages. By embracing many of these new techniques and guidelines, the proper operation of simulation models can be more easily ensured. Further, evaluation methods can be made more powerful and reassuring, leading to a more accurate assessment of the credibility of models and simulations.

In general, the Task Force has identified four areas of concern regarding modeling and simulation credibility:

a) Preserving the credibility of specialized, reusable model elements,

b) Evaluating the credibility of large-scale simulations used in the acquisition decision process,

c) Facilitating assessments of credibility by proper reporting of M/S methodology in appropriate DAB documentation, and

d) Exploiting new technologies for improving the credibility of simulations.

Summary of Findings

Modeling and simulation can be an effective tool in the acquisition process, if used throughout the system life cycle. Modeling and simulation contribute knowledge and understanding of system and mission requirements, system effectiveness, and costs resulting from acquisition decisions. The Task Force found that increased M/S effort in the following broad acquisition areas could be helpful to modeling and simulation concerns:

Mission Requirement Generation

M/S is used to establish mission requirements because it is generally the only tool available in the genesis of a weapon system acquisition, Figure 2-3. The use of M/S to establish mission requirements varies from program to program and from Service to Service. Two basic forces drive changes in mission requirements. The first is the push of advancing technology that make more effective weapons possible, and the second is the pull of the evolving threats. Both of these forces drive changes in mission requirements. The use of M/S to forecast the threat ten to fifteen years in the future is very difficult and riddled with vagueness. Also, the use of M/S to predict technological advances, although based on years of extensive research, is often quite imprecise. On the other hand, rarely do we predict the appearance of the truly revolutionary system (i.e., atom bomb).

To effectively set mission requirements it is necessary to model more than the weapon system capability of interest as shown in Figure 2-3. It is often necessary to model many-on- many systems, force-on-force engagements, and sometimes nation-versus-nation engagements. Modeling greater groupings of forces results in increased levels of aggregation with resulting losses in precision and certainty. Once the mission requirements are set the models are often set by the wayside until the next technology push or threat pull appears.

Proper M/S use can provide more illumination of choices in the operational requirement process. In the beginning of the acquisition process operational requirements are driven by advancing technology and evolving threat. M/S is the only tool available to evaluate the full ambit of force structures, doctrines, weapon systems, and threats to guide weapon system procurement. The requirements process is by nature inexact because it is difficult to forecast the threat, technological advancement, doctrinal changes, and costs ten to fifteen years in the future. A good requirements evaluation must include excursion analysis in many dimensions to properly account for and bound the uncertainties in the assumptions. It is also important that these models reflect reality as much as possible. For example, to properly investigate doctrinal changes it may be necessary to introduce people in the loop of the simulations. The M/S developed in the conceptual stage of an acquisition should be viewed as a tool that will be further developed and expanded throughout the development cycle.

Development Design Tool

M/S is used most extensively as a development design tool. M/S design tools often mature throughout the development process, often supplemented by development test results. M/S is used early in the development process for initial performance evaluation and first order system sizing. Later in the development process M/S is used for detailed system sizing and design verification. Near the end of a development program M/S is used to supplement the development test program to analyze regions of the envelope that are either impossible, costly, or unsafe to investigate through testing. M/S is used very effectively as a design development tool by the developing contractors.

Proper use of M/S can provide more flexibility in the development part of the acquisition process. The development portion of the acquisition process is often viewed as too rigid because of strict design requirements embodied in the technical specifications. The underlying goal of weapons system procurement is to acquire and field effective and supportable systems. The M/S tools developed in the operation

IMPROVING T&E EFFECTIVENESS

Role of M/S in Setting Mission Requirements

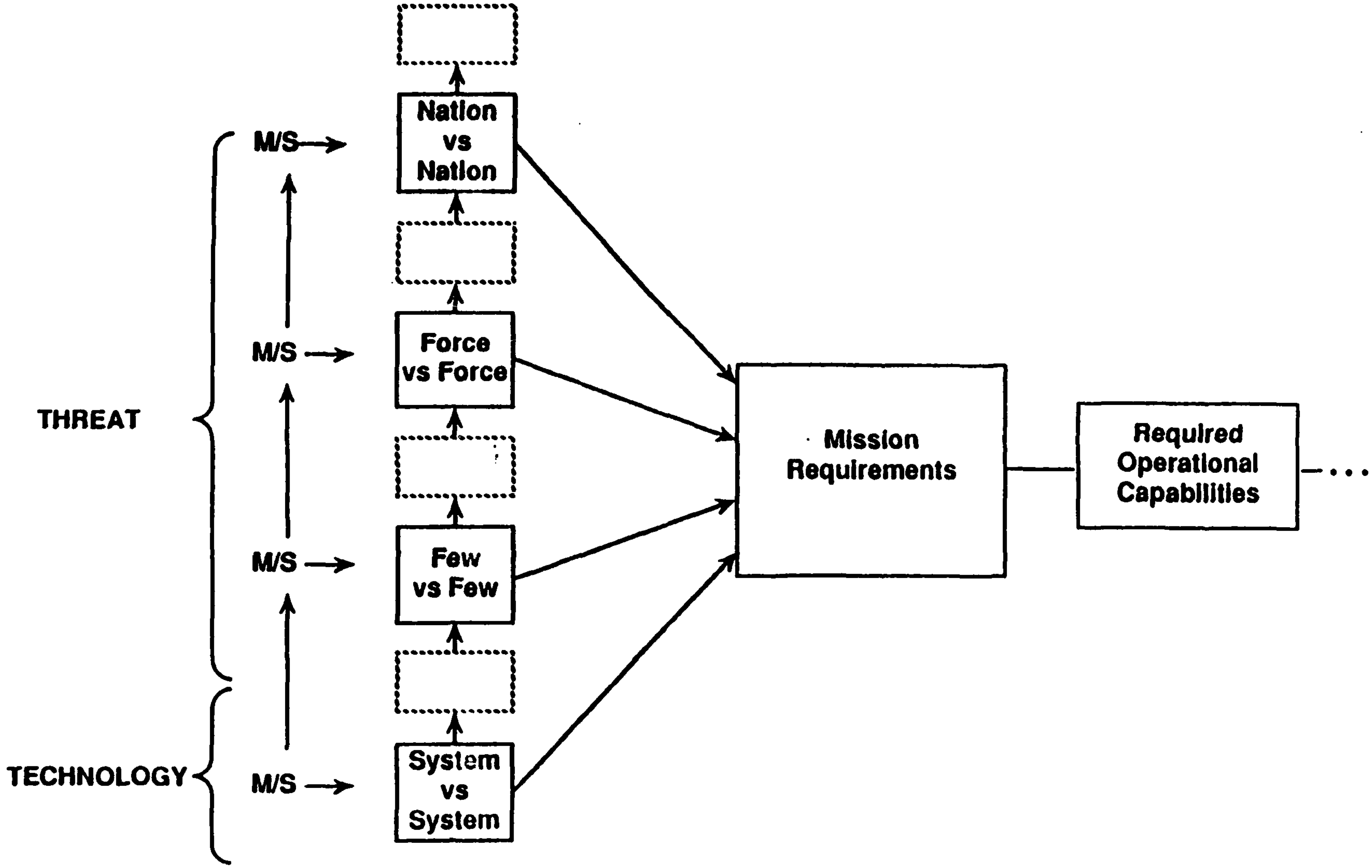


Figure 2-3

requirements phase could be utilized to evaluate the relative applicability of the technical specification throughout the development cycle, which is often five to ten years. This would allow revision of a technical specification in the development phase when the revision will yield a more effective or more supportable weapon system. This should not be viewed as a license to continually revise the technical specification, but as a Insert mechanism to revise a technical specification that originally may have been overly rigid. The M/S tools developed for operation requirement generation will need to be expanded and refined to accomplish these goals.

Cost Forecasting

M/S is used throughout the acquisition cycle to forecast weapon system costs. The credibility of these models has historically been suspect, as is evident in the press. There are some basic reasons for the lack of credibility in cost prediction models. Cost models are reliant on a vague set of assumptions in the beginning of the process. Early in the acquisition cycle there is often no clear description of the weapon system and definitely no design details. This is compounded by vast uncertainties in the value of the dollar ten to fifteen years downstream. Most weapon system developments involve some degree of technical risk that further complicates development cost predictions. All these factors and more contribute to making cost forecasting a very difficult task, analogous to forecasting the threat. Frequently, the office of management and budget dictated planning factors for cost escalation that will out weigh all other elements of cost factors.

It is evident that M/S plays a large role in the acquisition process but the Task Force believes that its use could be improved to further enhance weapon system procurement. M/S could be more effectively applied to provide more illumination early in the acquisition process, to provide more flexibility in the middle of the acquisition process, and to provide more consistent utility evaluation throughout the entire acquisition process.

Proper use of M/S can be used to provide a more consistent utility evaluation throughout the entire acquisition cycle. Often, the operational utility of an acquisition program is evaluated twice in it's acquisition cycle. The early evaluation is during the mission requirement and required operational capabilities section of the acquisition process. The final evaluation is during the operational test section of the acquisition process. The length of time between the former and latter is normally between six to ten years. A lot can change during that time span including evolving threats, new technology, and changing economic factors that could change the utility of the weapon system in development. A consistent evaluation (i.e., use of the same M/S tools and or tests) of the operational utility (i.e., effectiveness and supportability) throughout the development cycle would enlighten key decision makers during the budgeting process. This consistent evaluation process could be used to revise technical specifications when appropriate (as stated above) or to curtail or cancel development programs that provide little gains in operational utility. If M/S can shorten the acquistion process the enemy is denied the opportunity to field counter measures inside our production cycle.

SECTION 3 - CONCLUSIONS

Based on these findings, the Task Force came to the following conclusions:

a. On the role of models and simulations in the evaluation process;

1. The foundation of the acquisition process, the operational requirement and its translation into system terms, can be improved.

2. Early and continual involvement of the OT&E community can contribute to and improve the acquisition process. Modeling and simulation can assist in this effort.

3. A development program, as embodied in a specification and contract, may become overly rigid, restricting the willingness to evaluate and incorporate changes as threat, technology and knowledge evolve. Modeling and simulation can be used as a tool for continual evaluation of potential changes.

4. Accounting for human performance early in systems acquisition improves system capability and enhances the ability of the test and evaluation process to predict operational performance.

b. On the credibility of models and simulations;

1. Simulations used by DoD have proliferated. Frequently they are incomplete, too large or not thoroughly understood. Their credibility is questioned.

   Availability of high quality, reusable M/S elements could decrease redundant efforts while improving quality of key elements.

2. A process does not exist to provide an independent, objective evaluation of M/S utilization.

   The credibility of a M/S cannot be considered separately from its application to a specific problem, the validity of data inputs, and the qualifications of those executing the model and interpreting the results.

3. Current DAB documentation does not address model and simulation credibility.

4. Advancing technology in both hardware and software offer opportunities for improving the credibility and applicability of M/S. Continued research is needed.

SECTION 4 - RECOMMENDATIONS

Based on this evaluation, the following recommendations are made:

a. For the role of models and simulations in the evaluation process;

1. The Service acquisition executives should ensure that M/S excursion analyses are applied systematically to help reach and maintain agreement on major aspects of requirements and system performance.

2. The DOT&E and service OT&E communities should be chartered to participate early in the requirements process. In particular, they should translate operational requirements into an evaluation framework and document the roles for M/S as well as for testing in meeting evaluation objectives at each milestone and as appropriate between milestones.

3. The USD(A) should establish policy and provide guidance to the acquisition community for systematically reevaluating system specifications using M/S and test results.

4. Service acquisition executives should ensure that the development programs employ man-in-the-loop simulation beginning with requirements definition and mature the simulation along with the hardware throughout the acquisition process.

   JCS/CINCS should exploit technology capabilities in distributed computing and networking to simulate coordinated combined arms engagements with man-in-the-loop simulations and to evaluate results against live exercises.

b. For the credibility of models and simulations;

1. USD(A) should ensure refinement, maintenance and availability of models, weapon and threat data descriptions, and simulation elements having wide DoD utility.

   - Appropriate JCS and OSD offices should select/fund executive agents to maintain element repositories (DNA-Nuclear models, DIA-Threat data, JTCGs, etc.) complete with databases, code libraries, and documented limitations.

2. USD(A) should charter DDR&E to enable, as necessary, independent panels of experts to assess specific applications of M/S results on which acquisition decisions are based. The work would be tasked on a case-by-case basis and include participants from academia, industry, and the government.

3. USD(A) should modify DODI 5000.2 to require that DAB documentation (SCP/DCP, TEMP, COEA and CAIG) address the applicability of models and simulation. For example, the documentation could consider the M/S plan and methodology, limitations, assumptions, extrapolations, sensitivities, results, analysis, and validation.

4. DDR&E should continue to fund M/S technology at both the fundamental and application levels, including the M/S interfaces and languages, executable specifications, model interoperability, validation techniques and tools, and parallel and networked simulations.

APPENDIX A - TERMS OF REFERENCE

THE UNDER SECRETARY OF DEFENSE

WASHINGTON, DC 20301

30 MAR 1989

ACQUISITION

MEMORANDUM FOR CHAIRMAN, DEFENSE SCIENCE BOARD

SUBJECT: Terms of Reference - Defense Science Board (DSB) 1989 Summer Study Task Force on Improving Test and Evaluation Effectiveness

You are requested to form a Summer Study Task Force to examine the contributions of modeling and simulation to Defense test and evaluation so as to improve the acquisition process. Collateral tasks should examine ways to increase credibility, realism and objectivity in the test and evaluation process. This study should delineate the benefit to be derived from the timely use and role of validated models and simulations in lieu of threat targets and environments due to practicality, security, international treaties, and/or civilian encroachment considerations. The expected outcome should be test and evaluation initiatives required to definitize the scope and fidelity of testing required to: (1) support the quality production of defense systems; (2) evaluate and reduce the uncertainties associated with defense system acquisition/production decisions; and (3) support, evaluate and reduce the risks associated with introducing new technologies into defense weapon systems acquisition/production decisions.

The perception persists that less than credible, realistic and objective test results are obtained through the use of models and simulations in lieu of threat targets and environments. At the same time, system costs and real world test restrictions have increased our dependence on models and simulations. Task Force efforts should include:

- Review prevailing use of models, laboratory tests and field tests.

- Determine appropriate situations in which to test and/or model.

- What fidelity of representation of the system under test and its environment is required?

- What discipline governs interpretation of results and its application?

This Task Force will be sponsored by the Office of the Deputy Director Defense Research and Engineering for Test and Evaluation (ODDDRE(T&E)). Brigadier General Robert A. Duffy, USAF (Ret.) has agreed to serve as Task Force Chairman. The Executive Secretary will be Colonel Matthew M. McGuire, USA, ODDDRE(T&E/WSA), and the DSB Secretariat Representative will be Commander George A. Mikolai, USN. It is not anticipated that your inquiry will need to go into any "particular matters" within the meaning of Section 208 of Title 18, United States Code.

The Terms of Reference for this Task Force include no assignments to the Task Force that would indicate the Task Force would be participating personally and substantially in the conduct of any specific procurement or place any member in the position of acting as a "procurement official".

OGC: ________

Date: 12 May 89

APPENDIX B - MEMBERSHIP

T&E MODEL AND SIMULATION EFFECTIVENESS TASK FORCE

CHAIRMAN:
BGen Robert A. Duffy, USAF (Ret.)

MEMBERS:
LGen James A. Abrahamson, USAF (Ret.)
Mr. Norman R. Augustine
Dr. Solomon J. Buchsbaum
Dr. Raymond S. Colladay
Mr. Vincent N. Cook
Prof. Katherine T. Faber
Mr. David R. Heebner
Dr. Dennis R. Horn
Dr. Barry M. Horowitz
MGen Ralph H. Jacobson, USAF (Ret.)
Dr. Anita K. Jones
Dr. Paul G. Kaminski
Adm. Isaac C. Kidd Jr., USN (Ret.)
Mr. Sol Love
Dr. A. Louis Medin
Dr. John M. Palms
Dr. Victor H. Reis
Prof. Thomas F. Rosenbaum
LGen Philip D. Shutler, USMC (Ret.)
Gen. Donn A. Starry, USA (Ret.)
Gen. John W. Vessey Jr., USA (Ret.)

GOVERNMENT ADVISORS:
DDDRE(T&E): Dr. H. Steven Kimmel
DOT&E: Dr. Patricia Sanders
PAE: Dr. David Gallagher
DIA: Mr. Nick Bennett
DNA: VAdm. John T. Parker
Dr. Leon Wittwer
DCA: Dr. Robert Lyons
SDIO: Dr. Richard Bleach
ARMY: Mr. Arend (Pete) H. Reid
AIR FORCE: Mr. Carroll G. Jones
LtCol. Robert J. Schwarz

EXECUTIVE SECRETARY:
Colonel Matthew M. McGuire, USA

DSB MILITARY ASSISTANT:
Colonel Oliver Westry, USA

TASK FORCE SUPPORT PERSONNEL

Institute for Defense Analyses

Dr. Gary C. Comfort
Dr. Robert H. Boling
Dr. Mark C. Zabek

Center for Naval Analysis

Dr. Ralph Passarelli

Science Application International Corporation

Mr. Howard J. Harvey
Mr. Ramon L. Strauss
Mr. Edward P. Petkus

APPENDIX C

BRIEFINGS PRESENTED TO THE DSB TASK FORCE ON IMPROVING T&E EFFECTIVENESS

Meeting - April 11, 1989

Title	Presenter/Organization
Overview of Task Force Requirements	Kimmel/OUSDRE (DDTE)
Industry Perspective on Improving Test and Evaluation Effectiveness	Gansler/TASC
JCS Perspective on Test and Evaluation Modeling and Simulation	Roske
SDI Program Strategy	Bleach/SDIO/OSD
DOT&E Modeling and Simulation Concerns	Sanders/DOT&E

Meeting - June 6/7, 1989

Title	Presenter/Organization
Deep Fires Requirement & Threat	Reid/AMSAA
MLRS-TGW Program Overview	Reed/HQDA
Accredation Process for M&S	Beavers/AMSAA
COEA Methodology Overview	Jones/TRADOC
Countermeasures Modeling/Analysis	Palamo/LABCOM
Engineering Model Validation	Bradas/MICOM
6-DOF Submunition Trajectory Model	Sanders/MICOM
Drop Tests & Comparison to Models	Sanders/MICOM
Hardware-In-The-Loop Simulation	Cole/MICOM
Captive Flight Tests & Data Analysis	Bradas/MICOM
Battlefield Environment Simulation	Alongi/MICOM
MLRS-TGW Effectiveness Modeling	McClung/PMO
Reliability Growth Modeling	Foulkes/AMSAA
Test and Evaluation Master Plan	Foulkes/AMSAA
Software Development/IV&V	Holeman/SD,INC
System Trainers	Nally/Ft. Sill
Maintenance Trainers	Blount/OMMCS
Vulnerability/Lethality Method	Kirk/BRL
Modeling in Live Fire Test Process	O'Bryon/DDDRE
BFVS Modeling/Testing Assessment	O'Bryon/DDDRE
AMSAA's Role in T&E	Reid/AMSAA
Technical IEP/TDP	King/AMSAA
AMSAA System Perf. Eval. Method	King/AMSAA
Supportability Analysis Method	Morton/AMSAA
OTEA's Role in T&E	Dubin/OTEA
Modeling and Simulation in Support of OT&E	Dubin/OTEA

BRIEFINGS PRESENTED TO THE DSB TASK FORCE
ON IMPROVING T&E EFFECTIVENESS

Meeting - July 11-13, 1989

Title	Presenter/Organization
Use of M&S in OT&E	Dubin/OTEA
System Effectiveness Modeling	Croteau/PA&E
OSD Cost Modeling	Lee/PA&E
Use of M&S in Army estimates	Young, CEAC
Fleet Ballistic Missile	Fisher/FBMO
M&S concept paper presentation	Horowitz
M&S in Support of the RDT&E	Keil
V-22 Operational Requirement	Schaefer/USMC
V-22 Preliminary Design	Martin/Bell
V-22 Risk Reduction/Design Devel.	Taylor
V-22 Design/Test Support	Gaffey
Avionics/Flight Controls	Ballauer
Facility/Manufacturing	Hays
Threat and Vulnerability Analysis	Johnson
Reliability and Maintain Predictions	Monje
Operational Availability Models	Venzlowsky
Cost Analysis Models	Weathersbee
COEA M&S Support	Kusek
Flight Training and Simulators	Curtis
USAF DSCS Program Overview	Hartigan/DCA
Threat	Hartigan/DCA
Sys. Sim.-Integrated Test Facility	Groener/Army
M&S for Ground Segment Hardening	Miletta/Army
Nuclear Environment & Threat Sim.	Wittwer/DNA
Link Effects Evaluation & Mitigation	Bogusch/DNA
Performance Assessment & Validation	Bogusch/DNA
DSCS End-to-End Performance Model	Sims/DCA

APPENDIX D - GLOSSARY OF TERMS

AMRAAM - Advanced Medium Range Air-to-Air Missile

CAIG - Cost Analysis Improvement Group

COEA - Cost and Operational Effectiveness Analysis

DAB - Defense Acquisition Board

DIA - Defense Intelligence Agency

DNA - Defense Nuclear Agency

DOT&E - Director, Operational Test and Evaluation

DSB - Defense Science Board

DT&E - Development Test & Evaluation

ECM - Electronic Counter Measures

JCS/CINCS - Joint Chiefs of Staff/Commanders in Chief

FSD - Full Scale Development

MLRS/TGW - Multiple Launch Rocket System/Terminally Guided Weapon

M/S - Modeling and Simulation

ODDDRE(T&E) - Office of Deputy Director for Defense Research and Engineering (Test and Evaluation)

OSD - Office of the Secretary of Defense

OTA - Operational Test Agency

OT&E - Operational Test & Evaluation

PM - Program Manager

RFP - Request for Proposal

ROC - Required Operational Capabilities

SCP/DCP - System Concept Paper/Decision Coordinating Paper

SPO - System Program Office

TEMP - Test Evaluation and Master Plan

USD(A) - Under Secretary of Defense Acquisition

APPENDIX E
TASK FORCE BRIEFING
IMPROVING T&E EFFECTIVENESS

TERMS OF REFERENCE MANDATE

TO EXAMINE CONTRIBUTIONS OF

MODELING & SIMULATION

TO

DEFENSE TEST & EVALUATION

TO

IMPROVE THE ACQUISITION

PROCESS

T&E MODEL AND SIMULATION EFFECTIVENESS TASK FORCE

CHAIRMAN

BGen Robert A. Duffy, USAF (Ret.)

MEMBERS

LGen James A. Abrahamson, USAF (Ret.)
Mr. Norman R. Augustine
Dr. Solomon J. Buchsbaum
Dr. Raymond S. Colladay
Mr. Vincent N. Cook
Prof. Katherine T. Faber
Mr. David R. Heebner
Dr. Dennis R. Horn
Dr. Barry M. Horowitz
MGen Ralph H. Jacobson, USAF (Ret.)
Dr. Anita K. Jones
Dr. Paul G. Kaminski
Adm. Isaac C. Kidd Jr., USN (Ret.)
Mr. Sol Love
Dr. A. Louis Medin
Dr. John M. Palms
Dr. Victor H. Reis
Prof. Thomas F. Rosenbaum
LGen Philip D. Shutler, USMC (Ret.)
Gen. Donn A. Starry, USA (Ret.)
Gen. John W. Vessey Jr., USA (Ret.)

T&E MODEL AND SIMULATION EFFECTIVENESS TASK FORCE (CONTINUED)

GOVERNMENT ADVISORS

DDDRE(T&E): Dr. H. Steven Kimmel
DOT&E: Dr. Patricia Sanders
PAE: Dr. David Gallagher
DIA: Mr. Nick Bennett
DNA: VAdm.John T. Parker
Dr. Leon Wittwer
DCA: Dr. Robert Lyons
SDIO: Dr. Richard Bleach
ARMY: Mr. Arend (Pete) H. Reid
AIR FORCE: Mr. Carroll G. Jones
LtCol. Robert J. Schwarz

EXECUTIVE SECRETARY:
Colonel Matthew M. McGuire, USA

DSB MILITARY ASSISTANT:
Colonel Oliver Westry, USA

TASK FORCE SUPPORT PERSONNEL

Institute for Defense Analyses
Dr. Gary C. Comfort
Dr. Robert H. Boling
Dr. Mark C. Zabek

Center for Naval Analysis
Dr. Ralph Passarelli

Science Applications International Corp.
Mr. Howard J. Harvey
Mr. Ramon L. Strauss
Mr. Edward P. Petkus

Outline

- Acquisition Process
- Role of T&E in the Acquisition Process
- Role of M/S in the Evaluation Process
- Credibility of M/S

Terms Of Reference Mandate

To Examine Contributions of Modeling and Simulation

To

Defense Test and Evaluation

To

Improve the Acquisition Process

[Slide 1] Our panel was asked to look at modeling and simulation as they pertain to test and acquisition of defense systems. In the course of doing our study, we discovered that in order to make useful suggestions on modeling and simulation we had to make corresponding recommendations in the areas of test and acquisition. So we broadened the scope of our activity to include all of these processes. That makes the subject of our study pretty complex, so I am going to begin the briefing by providing an overview of the thought process used for relating all of these processes. This should provide the context for the more detailed discussion with regard to the study.

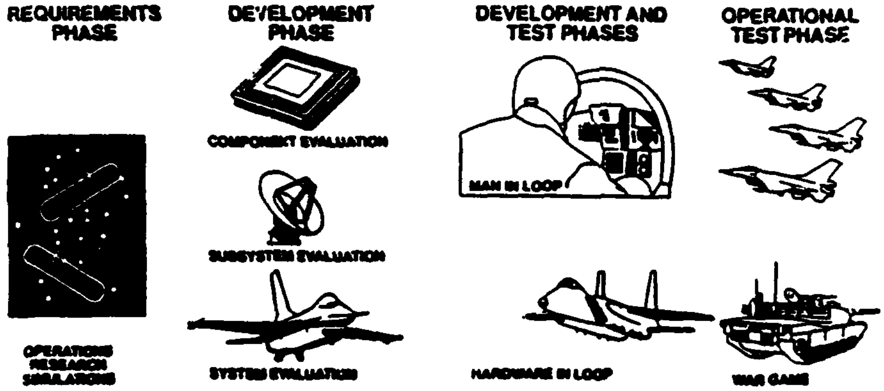
We recognized that over the last ten years there has been a continually-increased emphasis placed on the use of special government test organizations, independent of the acquisition organizations, to define, conduct, and evaluate the results of operational tests on newly developed systems. This has been done to improve the acquisition process by adding confidence to the production decisions; those buy, no-buy decisions that weigh operational test data heavily. A significant side effect of this emphasis on the use of operational testing as an audit tool after development, has been a corresponding shift in the outlook of the development community on the overall role of operational testing. This change has been to de-emphasize the use of operational testing as a learning tool during development . That is, something the developer does while designing a system to better understanding the complicated interrelationship between the specifications for a system, the design for a system, and the ultimate operational utility that is being sought. Why does this correlation occur? Well, look at it through the eyes of a program manager who sees the world as ... his job is simply to develop a system. A user substantiates the utility of the system both before and during its development, and an independent organization evaluates its utility once it is developed. So, the program manager's role is simply to be a provider of the system with no specific requirement to deal with the utility of the system.

So, the question arises, "What are the consequences of this deemphasis on operational testing from being a learning tool during development." We will show a number of examples to indicate that the consequences are negative, and they are serious enough that we make a number of recommendations to increase the use of operational testing during development as a learning tool.

What does this have to do with modeling and simulation? Let me defer that for a minute and provide some background on modeling and simulation. A model is an idealistic representation of a system. A simulation is a process that permits us to evaluate a system by stimulating a model with assumptions-inputs-and observing responses, - the outputs.

DSB IMPROVING T&E EFFECTIVENESS

Use Of Simulation In The Acquisition Process

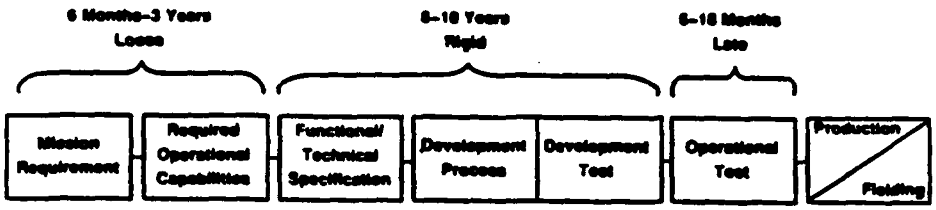


[Slide 2] Of course the Defense Department has been using simulation for years and years and in many different ways ranging from, operations research type simulations which help in the very early planning stages of a new product, to understand the utility, to simulations used by designers and developers, such as microelectronics designers or radar designers, to simulations to deal with human factors, such as cockpit simulators., simulators that synthesize complicated environments that real systems will operate in, such as electronic combat environments, and ultimately the war-gaming simulations which are usually the basis for our operational tests. Over the last ten years we have seen the dynamic growth in the computing and networking technology areas which are the underpinning for digital simulation. So we have seen a corresponding increase in the use of simulation. Simulations that can be much more elaborate and much lower in cost than they were in earlier times. It is not a very risky prediction to state that this will probably continue to be true for the next ten years. Since there aren't many things that get cheaper and better with time, a question arises as to how we can take advantage of this evolution of technology to permit simulation to have more value. Now back to my discussion on operational testing during development for its learning value. It is here that we are going to focus our attention on simulation.

First, we recognize that operational testing during development is expensive, and no one is going to do it without clear value. So the question then is, "How do we know what tests are worth running?" It is that role that we are going to assign to simulation; namely, conducting simulations to help us identify and focus on areas of concern related to the ultimate utility in operations of a product that is being developed. When the concern is sufficient we should be willing to run operational tests targeted at the issue raised by simulation. How you do this will be discussed later in the briefing.

Another question arises about simulation. We have all used simulations and had experiences with simulations which gave us wrong results. Somehow or other, no matter how many good simulations we have worked with, everyone always remembers the ones that gave poor or misleading results. So there is a natural hesitancy to believe simulations. This is a good thing, but the question arises as to if we are going to make more use of simulations, what do we do about this credibility factor, especially when related to the decision process in the Defense Department. While I talked earlier about production decisions and operational testing, there are many other very high-value decisions made prior to production decisions that don't have operational tests as their basis, but rely more on simulation. So the question arises as to how does the decision-maker know whether he's making a decision based on one of these bad simulations or a good one? This question has been asked in many forms to our panel: "Should a central organization be created in the government to accredit simulations that are permitted to be used in decision making processes similar to what has been done with operational testing -- bring in an independent team for greater confidence?" "Should the government set up a central management organization to not only accredit simulations as being good ones for use in important decisions, but to manage the distribution and re-use of those standard simulations across a wider segment than might otherwise use them?" Over the course of the briefing I will bring in this issue of credibility. With that, I am going to get into the details of our study.

Acquisition Process



[Slide 3] I am going to begin with a discussion of the acquisition process, the first of those three processes that we are going to try to interrelate. We have a very simple model here of an acquisition process. It begins on the far left and has three parts: the beginning, the middle, and the end. They are what we usually think of as the requirements, the development, and operational test phase for production decision making. If we look at the beginning, that is, the period where people in the services generally are desirous of improving their capability to perform a given mission and are imagining how some new system or new technology can be exploited to make that possible. To do that, they have to somehow or other extrapolate their past experience in military operations with their future vision of systems and technology to bring about some vision of how a new capability can make them more capable. This involves making assumptions such as what the enemy threat will be, what the doctrine will be for our own and for enemy forces, what the procedures for using this new capability wil be, what the deployment of the new system will be, etc.. So a number of predictions have to be made that go along with the extrapolation of the utility. During this period people get a stronger or weaker feeling about the need for a system and sometimes we use those operations research type simulations I referred to earlier to give some substantiation to the level of utility before proceeding with a new development. In the sense that it requires lots of prediction and extrapolation, the process, by its nature, is termed loose. It is not that people don't do a good job or that they are not doing their best, it is just that by its very nature this process has to be loose; involving predictions about our own forces as well as enemy forces. When people get a strong enough feeling to go on with development, they then bridge into that middle phase, the development phase. The method of bridging is the creation of a technical specification for the product to be developed. That technical specification is created by two groups: an operational group, who had that vision in the beginning, and the group that is going to be responsible for development, which is a more technically-oriented group. Through a process of further extrapolation, that is extrapolations about lower levels of details on the technology and lower levels of detail about what might be the coupling between the specification and the operational utility, a specification is born. By its very nature it is an erroneous process. Again, not errors due to people making blunders, but the specification generation process involves even greater levels of prediction and extrapolation than the loose process upon which it was founded to start with. Once that specification is created, it becomes the basis for a contract, a contract that must be rigorously managed. So we shift into a mode from loose things and error-prone things to something that has to be managed rigorously. There is nothing wrong with that; it's a necessity of contracts. If one looks at the usual length of the development process as this chart shows, it lasts many, many years. The program manager who does a good job is one who keeps his programs stable through rigorous management.

We call this phase "rigid" in the sense that we go from a foundation which is loose, to a management process which is very rigorous and which is founded on keeping things stable. We then bridge into the final phase which is the operational testing.

Here an independent group is brought in to determine whether or not this system really has enough utility that it is worth producing. There are really three evaluations going on in this phase. One is the evaluation of the equipment itself; the second is the evaluation of the very early work done by the planners who imagined what the utility would be if such a system were built. The third is the evaluation of the process that translated the vision of those operational planners into a specification that was the basis for building the system, which itself is subject to error. So we are testing three things, two of which could have been tested ten years ago if we had the equipment. So, in the sense that two of the three things have had a very long period of delay without much additional work, these tests are very late. They are also late in the sense that the cost of discovering a problem in this stage of acquisition is at its highest. Either the cost of the development is sunk, if a program is cancelled, or if a decision is made to rehabilitate the system, the cost of rehabilitation is at its highest because in addition to doing the redesign work for the product, all of the support costs associated with changing drawings, changing support equipment, etc., in keeping with the modified design must be borne. When we, in fact, find a failure in operational testing, we will refer to that as a surprise because it is hard to imagine that somebody would go on developing a system for ten years in anticipation of a failure at the end. As everyone knows ,in the Defense Department, surprises are very bad, because in addition to the surprise of the details of that particular development, the credibility of the whole acquisition process is brought to the surface. And when the process itself has a lack of credibility, it has a side effect on everything we do that is negative and inefficient. So it is really very important, from our panel's point of view, to avoid surprises both for the instantaneous cost on the system in question, and the credibility loss to the process as a whole.

Two Study Issues

- **What can be done to help avoid surprises during independent operational testing?**

[Slide 4] So that raises the question, "Can we do anything to avoid surprises?" There are two presumptions here. One is that we have them and two, that hopefully they are avoidable. So let us talk about whether we have them or not.

Typical Surprises

- "Change in Assumptions" Surprise

[Slide 5] The first kind of surprise I am going to discuss is what we call "the change in assumption" surprise. This is where those early planners made some assumptions which led them believe that the system or technology they wanted to advance would be useful, like the threat, the deployment of the system if it were going to be built, the environment it would operate in if it were used, and on and on. Then ten years of development go by and for whatever reasons the threat changes, the deployment plans change, the key features change, some of the basic assumptions change. Then the operational test is run and it is determined that those changes are so crucial that the utility of the product goes from being something that was imagined to be useful, to something that doesn't seem like it is worth it. A good example of this is the DIVAD. The DIVAD was an Army air defense system. It was originally conceived in the early '70s and development started in 1977. Its purpose was to protect the moving army from close air support attacks by fixed-wing aircraft and from standoff helicopter attack. At the time the program was initiated, the standoff helicopter threat was seen as a three kilometer range standoff weapon. So the designers of the DIVAD decided it needed a four-kilometer firing range. Well the development went on through 1985, and during that period of development two things happened to that threat. One is the helicopter threat became primary, and second, the range of that firing capability of the helicopter increased to six kilometers. Well, the operational test was run and it was determined that the firing range of the DIVAD was inadequate given the extended standoff range of the helicopter threat. The result was that the program was cancelled after $1.2 billion of investment. Now it is not as if the development community did not know that the threat was changing. They did, but they had lots of rationale as to why it was still logical to develop the DIVAD with its four kilometer firing range. The point I want to make is not that we should have produced, the DIVAD, but why did we have to wait until the end of the program and having spent $1.2 billion to decide that the change in assumptions was crucial. Is there something we can do to make this sort of thing occur earlier?

Typical Surprises

- "Change in Assumptions" Surprise
- "Measure of Effectiveness" Surprise

[Slide 6] The second type of surprise I want to talk about is what we call the measures of effectiveness surprise. This is where the early planners had some way that they thought about utility. If the system could do it good, that would be useful. Then we go through the development process and an independent team comes in to run the test and they come in with a different set of measures. Different enough that what might seem like a very useful system no longer appears to be useful. The result being that we do not produce the system. An easy generic case for thinking about this is when one person says, "If the system under evaluation results in a capability that is a lot better than is in the field today, and I see no other way of getting it in the near future, that is good enough for me." Whereas another person says, "No, no, no, I draw a chalk line and unless you're better than that, the system is not useful at all." A good example of that is the Aquila, another Army development. The Aquila was an unmanned airborne vehicle and its purpose was to carry sensors to provide a capability to take advantage of the extended firing range of artillery that the Army already owns. These weapons can fire to 15 or 20 kilometers beyond the FLOT; weapons such as the 155 Howitzer, and the MLRS rocket system. Yet they have no way of knowing what targets are at that distance unless they send out airborne spotters or ground-based spotters; not a very effective way for seeing what is out there. So the idea is to provide an unmanned airborne vehicle with a TV sensor, IR sensor, maybe even a laser designation system to aid those weapons. The original planners in 1974 said that, "...the system was useful if the sensors could see half of the targets out in their area of vision, and when targets were observed, 85 percent of the time the weapons could exploit it and actually kill the targets. If the vehicle was not that hard to use, (e.g., it would take less than an hour to get out on station and do a usable job), that set of capabilities would really be useful to the Army." Well this program went on from 1974 to 1987 and in 1987 an operational test was run. It only included the TV sensor. It did not include the full capability. In the conduct of the test there were some confusion factors, like people didn't know how to operate unmanned vehicles very well. They had little experience in doing that, but nonetheless the test was run and the system generally satisfied all of those original measures of effectiveness. Nonetheless it was determined that those were not good enough to warrant production and the system was cancelled. That was an $812 million dollar development. Now in looking at the report, there is no indication as to what was good enough. It was only stated that the system was not good enough. So that raises the question, "Should we go 13 years through a development with the designers and planners having a vision of what was good enough and then arrive at a decision point where it was decided that this was not good enough?" Why couldn't we have had substantiated measures that the community as a whole had adopted as the ones? I will also point out that many senior military people still believe that we should produce the Aquila and that, in fact, if you look at it, there is nothing coming down the pike to give that extended range to the capable weapons.

Typical Surprises

- "Change in Assumptions" Surprise
- "Measure of Effectiveness" Surprise
- "Lack of Maturity" Surprise

[Slide 7] The third type of surprise I want to talk about is what we call the "lack of maturity surprise." This is where a very natural tension in development always occurs. You get the system to the end of that middle stage of development and it's time to decide to start the operational testing. You have your very first units of systems available for test. They've had no time to mature. Certain parts of a design in a system need to have real use experience before they mature such as software bugs and hardware reliability, and so you have a decision to make. Should I take the time to get that experience, add that maturity, and therefore have a better chance of running an operational test with success, at the expense of leaving a factory idle; a factory that has a lot of people and a lot of machines ready to produce. On the other hand, should one enter the test, what might be considered as prematurely, in the sense of maturity, but hopefully be able to get through an operational test, get a decision to produce, get that factory working as quickly as possible, and do the maturing during the long lead time it takes to get the first production units out? Almost invariably the Defense Department takes the latter course of a riskier entry into operational tests to gain the economics of rapid production. This is probably a good thing to do; however, on occasion, you run a test where those maturity things bite you. A good example of that occurred with JTIDS, a tri-service airborne datalink system that was being evaluated. In this case, the objective was to have a system with 400 hours mean time between failures via ground test as a capability. When JTIDS entered the operational test it had about ten percent of that. As a result, the operational testers noted that the reliability didn't seem very good and, in fact, it disrupted the ability to run good operational tests. So they rightfully stated that the reliability appeared inadequate. Now this was no surprise to the development community. They knew that these units had not had a chance to mature, but it was a great surprise to all of the Tri-Service people who are presented with the results of an operational test, and who have no idea about the status of maturity when the test starts. This raises all kinds of issues like, does the organization developing the system know what it is doing? Is the system any good? Is the management group in the government incompetent? The subject transfers from a test of JTIDS to a test of the credibility of the whole acquisition system that created JTIDS. As I stated earlier, that really does not do anybody any good and it is also very inefficient. Why is it inefficient? Red teams, special panels, briefings to everybody who has any affiliation with the use of that system and investment to make, start at a very rapid pace and use up a very long period of time before credibility is regained. In fact, on the JTIDS program, what happened is over the period of a year or so, while all that was going on the real system was being matured and eventually showed via tests, about 80 percent of the mature reliability, and in fact, is now back in a more normal mode of development. But at the expense of a very long period of credibility loss; credibility that will probably never be fully regained in that system. Perhaps we should share the knowledge of the state of maturity before we enter operational test with the full set of involved players, rather than have a very large set of people be surprised by the end result of the demonstration, that the system is not mature.

Typical Surprises

- "Change in Assumptions" Surprise
- "Measure of Effectiveness" Surprise
- "Lack of Maturity" Surprise
- "Lack of Usability" Surprise

[Slide 8] The last example I want to give I call the "lack of usability surprise." This is where, for whatever reasons the user, the real user, the soldier or pilot or sailor who is going to use the system, has had no chance to try the system until the operational test. That is, the development community had surrogates trying it during development and so the very first trial of its use is during operational test. The surprise is when, in fact, the real users do not like it. They say the system might be good and all that, but we cannot use it. It is too hard to use. We just do not like it. We do not want it. This really can happen. A good example of this is a digital network built for the Strategic Air Command, called SACDIN. This was a network whose purpose was to disseminate emergency action messages to the strategic force structure and to receive status back from that force. As you can imagine, this has to be a very secure network. The system that was being developed, SACDIN, utilized message entry terminals comparable to your workstation of today that might exist in your office for a computer system, except these workstations had to assure security and there were all kinds of software measures taken to assure security. Well, the system was developed over the normal course of time and the operational test was run. One of the features built into this system, (and you have to understand this was the most secure system imagined for that time, nobody had ever gone this far in building security, so the designers for security were very nervous) was that if a terminal had a lot of wrong inputs put in, in sequence, it was possible that this was a security breach -- someone was trying in some way to disrupt the system -- and they froze the terminal, audited the most recent data and blew a whistle to bring in a security officer. That was how the system was designed. Now you get to the operational test and you have a user using it who does not have a lot of experience with that particular system and makes some errors, and he makes enough of them that, in fact, it satisfies the criteria for being a potential security breach and it freezes. At the end of the test the individual user says, "I cannot use that stuff, every time I make a mistake, instead of helping me it freezes, it is not useful, I do not want it." Well you go and fix that. You change some software; that usually handles matters like this and it is fixed. Well, to achieve the computer security, one had to have a specification that was mathematically verified as maintaining security transfers, manual validation that the real software matched the specification, and then employ professional teams whose job it is to try to do code breaking, to see if they can break the system or not. All that has to be done through a regression test process to accept the new changes in the software that satisfy that man-machine problem. In that case we spent about a year redoing SACDIN to solve the problem. In our panel's view, today this should never happen. You should never have a case where man-machine interaction has not been evaluated by use of rapid prototypes and simulation.

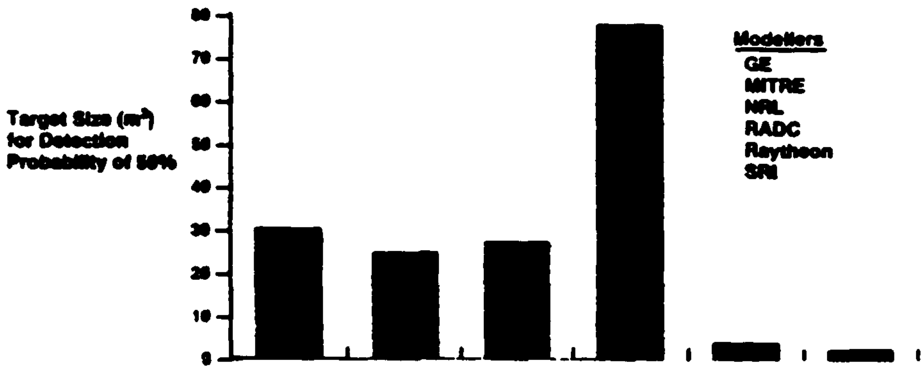
Two Study Issues (Concluded)

- *What can be done to help avoid surprises during independent operational testing?*
- *How can we increase the value of modeling and simulation in the acquisition process, and be confident of its use?*

[Slide 9] Now I am going to move away from tests and get into the area of simulation. Then later I will bring it all together with our panel's recommendations. So let me talk about simulations. I am going to begin by discussing that topic of confidence in simulation that I referred to earlier and then I will get into the role of simulation to help us focus on operational testing during development as a second subject. So let us talk about confidence.

Comparison Of Simulation Results Among Six OTH Radar Organizations

Results of simulations run in 1986, all on the same case:
AF OTH Radar in Maine, winter, 1,200 mile range,
azimuth = 70 degrees, nighttime, sun-spot number = 25



[Slide 10] This first set of examples, is to give an indication of how hard it is to assure confidence in simulations. The first example is associated with an over-the-horizon radar activity. Over-the-horizon radars, are high frequency radars, which transmit high frequency electromagnetic waves that bounce off the ionosphere and illuminate targets well beyond line-of-sight, at very long range. Then these signals are reflected back at the radar for targets via comparable return paths. These were developed to see 100 to 200 square meter size aircraft targets, such as bombers that might be attacking the United States. During the period of time in which these early systems were developed, questions started to arise about cruise missiles because the Soviet threat was changing to having standoff cruise missile capability, which had smaller cross sections. The question was asked, "What sort of a modification could be made to these radars we already invested in for bomber sized targets to make them capable of seeing cruise missiles?" Even further questions were asked like, "Do I have to modify them at all?" Because it turns out that if you know about HF propagation, you understand that the performance of the radars is very situation dependent. It depends on the path of propagation, it depends on the time of day, it depends on the time of year, it depends on where in the sunspot cycle the period is that you are operating in, all because of the effects of the ionosphere. So a lot of margin is put into the design of over-the-horizon radars to handle off normal situations. So it was not a ridiculous to say, "Well maybe in the more nominal propagation conditions you could actually see very small things even though the radar was designed to see bigger things." Well a bunch of experts, and they are listed on this chart, who have a long-term experience with over-the-horizon radar were going around giving different answers to this question. So some very wise person in the government said, "Well, you know, it could be that they are all thinking about different assumptions when they are answering the question, and HF propagation is so assumption dependent that what we have to do is get them on an even ground on the assumptions." So the government set up a special process where all of these people who had simulation models that they had used regularly in the course of dealing with their bomber sized target analysis and reliably so, were asked to look at identical circumstances and give answers to what size targets could these radars see under a given circumstance? Each of those black bars represents, for this one case that is shown on this chart, an answer that one of these organizations gave. And as you can see, they vary from one organization saying, "80 square meters is the size you can see," to others saying you could see targets in the ones of square meters, which is more in the range of what we are talking about when talking about cruise missiles. So how could it be that all these experts with validated simulation models, gave such widely different answers? Well, obviously the extrapolation of the models for this new question were unreliable, at least for a bunch of them if not all of them. So what do you do? Well, two things were done in reality. First, it was decided to run a real test, and drones were flown which were to be facsimiles of the real target to see what the radars we had were capable of doing. You would need to run lots and lots of experiments to get all the cases, but at least the tests would provide an ability to calibrate the simulation models. Now the second thing, you could look at the details of the models, which was done, and it turns out that they lacked fidelity relative to accounting for all the propogation phenomenology, such as ionospheric focusing, and multipath effects. Phenomenology effects which were not critical in dealing with 100 to 200 square meter targets, had become very important in dealing with smaller targets. So we have a case here of valid models extrapolated for use to problems that seem like they're the same problem, but are different enough that the extrapolation is erroneous. The point being, that if I want confidence in a simulation model, I not only have to know about the model, but I have to know about the extrapolation involved in dealing with the specific problem. So you have to know two things very well.

Marketplace Validated Models Example

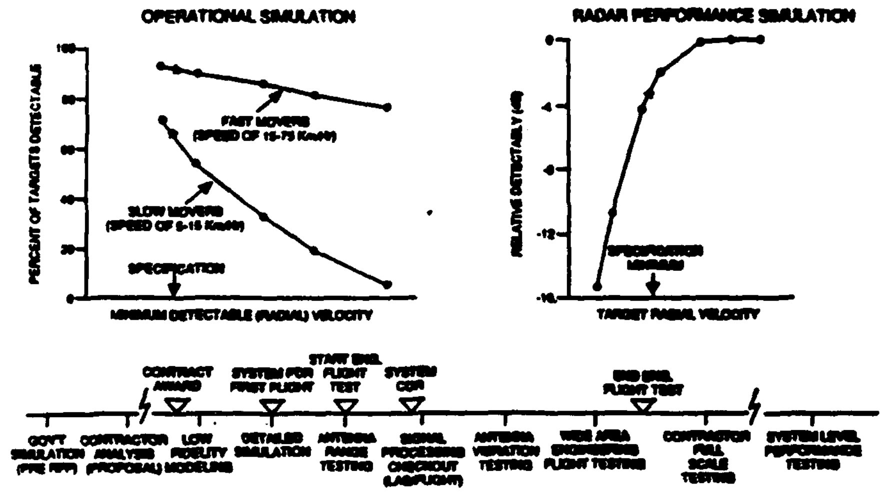
Response Time Models for Data Processing System	Electronics Reliability Models
SLAM II	RPP
STATEMATE	Oracle
AISM	Predictor
PAWS/GPSM	REAP
BEST I	Mistress
•	•
•	•
•	•
•	•

[Slide 11] The next example I want to deal with is what we call, "marketplace validated models." These are models which companies create, sell, and other companies use all the time. The government uses them all the time also. And the marketplace is the test of validation in that people buy them and use them. Again, I am going to illustrate problems with validation with valid models. The first example I want to give are these response time models for data processing systems. Here the question is, "I am building a big distributed data processing system, maybe a worldwide system with a lot of users on it, and I would like to know from the time a user pushes a button requesting data or asking for some function to be done by the system, what is the time it takes to get a response?" It is a very normal question to ask. It turns out that one of the key factors in determining that answer is when we call "contention"; that is, when other users happen to make requests that ask the same processors to function at the same time as this first user who I asked the question about. If there is a lot of contention and the details of how that is technically managed are inefficient, responses could take a very long time. And if there is not a lot of contention, and the details of the technical management is very efficient, responses might not take a long time. Well, an input to the simulation models then is a prediction about contention. Well, how does some software designer sitting in a factory designing software know what the contention will be in some system that has not been fielded yet? He doesn't. So he makes some judgment as to what he thinks the contention will be. People with experience in this field know that we very frequently make misjudgments on this, and as a result we get very wrong predictions out of these kinds of models. So this is a case of what we call "garbage-in-garbage-out". Now what do you do about that? Well, experienced people know you try to get earlier versions of a system out into the field and get some early measurement of what this contention might really be based on field experiments. In that way, they are able to provide some validation to the data and avoid garbage-in.

The other example has to do with reliability models. The problem here, is one we all have to deal with. Namely, we have the design of a piece of electronics sitting on a desk and we are curious about what its reliability will be before we go and develop it. These models will tell you that, and they will account for things like the quality of the parts, the thermal stresses, and the electrical stresses that the system will have in operation, and through some integrated set of calculations will determine what the mean time to failure will be. People use these all the time, but people with a lot of experience know that this only accounts for a set of factors that deal with parts factors. It does not deal with factors like workmanship, manufacturing quality, issues like the mechanical rigidity of the boards used, whether in fact they will buckle and the connectors that you have chosen, are adequate to deal with that sort of thing. So it really gives you an accurate answer for a part of the problem, but it is a part of the problem that determines only a fraction of the reliability. So people who know about this are wary about the complete answer, but nonetheless see the value of the partial answer. That is, they are assured that the parts will not take them down in terms of reliability. Somebody who is validating the model cannot really know what each factory's quality is and what the workmanship standards are in each factory. It gets to be a very specific determination. So that is another way that the ability to validate a model really does not validate the result when presented for decision making.

So the point of those three examples was to illustrate that this concept of validating simulations is really a very tricky business. It is difficult to imagine how one group could do that.

Joint STARS Example Of Simulation And Test Performance Verification



[Slide 12] Now I want to switch to the subject of operational testing and the role of simulation coupled to that. I am going to use an example which we believe is an outstanding example of how simulations should be used, but we feel it is not common enough in defense acquisition systems. This has to do with Joint STARS. Joint STARS is an airborne radar system that is capable of seeing slow-moving vehicles like tanks on the battlefield and serves as a surveillance resource for many weapons. Let us look at this chart. It is complicated, but I think it's worth the effort. The graph on the upper left is an operational simulation. What this simulation does, is it deals with the utility of the surveillance system. And the way it does this is as follows: first we recognize that the ability to sense moving targets in Joint STARS depends on sensing the velocity of the target directly at the radar beam, the radial portion of velocity. So if a target is moving at any speed, but orthogonal to the radar, it is not seen at all. If it is moving directly at the radar beam, it can be seen depending on what minimum velocity that radar is sensitive to. So the question is, "What should be the minimum velocity that radar is sensitive to in order to have utility... ten miles an hour, five miles an hour, a mile an hour, what should that be?" Well the curve on the upper left that deals with that subject is derived as follows: a model was created which took the roads of Europe and the off-road areas that were usable by tanks and trucks. The modelers laid a hypothetical Soviet force of moving vehicles down on this area, and they then calculated by geometry the fraction of the total speed of each of those vehicles that would be pointed at the radar's beam. They then computed if the radar could only see some minimum velocity or greater, what fraction of that target set would be visible? And what the curve shows is that the specification turns out to be ____ kilometers per hour. Then if we could see ____ kilometers per hour, we would see about 75 percent or 70 percent of all the slow-moving targets like tanks, and we would see 95 percent of the fast-moving targets. And if we could not see ____ kilometers per hour but could only see greater velocities, the fall-off rate is pretty fast in terms of the percentage of the tanks that this system could see. So from this curve, people became very desirous of having that radar sensitive to very slow speeds. You can understand that. Well a corresponding simulation was done which is shown on the upper right, and that was a simulation done by radar designers. They did a design of what sensitivity would this radar have to have in order to see targets of some minimum speed and faster. They did simulations which ended up with a design curve of the sort shown here, which showed that if we take the specification value, a given level of sensitivity would be needed to see ____ kilometers per hour. To see slower speeds, you would have to move down a very steep curve of sensitivity improvements to see even small increments of lesser velocity. They knew they were pushing the technology as hard as they could even to get to the specification value on the curve. And that, in fact, is what ended up determining where the specification should lie. Now they could have stopped there and in a briefing to decision-makers said, "Well, the spec is ____ kilometers per hour, we have done some simulation, and we meet the spec." But they did not do that. They also looked at the top part of the curve above the star and they recognized that if their simulations were inaccurate, by not so large an amount, they could drift into a portion of the design curve where they would be losing a large amount of sensitivity to slow-moving targets. So it becomes important, given that a drift in that direction due to a simulation error could yield a significant drift downward on the other curve of utility. So it is a case where you have two slippery slopes. And when you have two slippery slopes, it is a good rule to do added work to make sure that you do not slip.

Given the focus simulation put on a potential operational problem, an aggressive program to do early operational evaluation were established, to determine whether these simulations were accurate or not and so that they would not wait 10 years to discover whether or not this product has utility. The first part was to take the budget of input errors that creates that design curve: antenna design, antenna stability, oscillator stability, the algorithms for signal processing, lots of things; and understand the input that creates that output, and be sure that they were not doing garbage in-garbage out. Since you cannot have the system the first day of the program you can at least start with the inputs and in fact after doing the simulation work they have been doing antenna range testing, algorithm testing, vibration testing, all as precursors to understand that at least they had the right inputs to their model. They scheduled to run experiments at the earliest time possible, with subsets of the whole system, but as full a capability for the radar as possible. They have scheduled an operational test in Europe where they can determine that those design curves are, in fact, accurate; accurate enough that they are not going to slip down that utility curve. At the same time they are going to start to data link the data to the operational users so that they can start to get early indication that the system will have utility at the end point. It is this concept for simulation that we think should be stressed, namely, as a focusing mechanism for running expensive, but very useful operational tests, as early on during the development process as we can, rather than waiting those six to ten years for the independent test to be the first crack at understanding whether the systems we are building have utility or not.

T&E Model And Simulation Effectiveness Task Force

Chairman

Brig. Gen. Robert A. Duffy, USAF (Ret.)

Members

Lt. Gen. James A. Abrahamson, USAF (Ret.)
Mr. Norman R. Augustine
Dr. Solomon J. Buchsbaum
Dr. Raymond S. Colladay
Mr. Vincent N. Cook
Prof. Katherine T. Faber
Mr. David R. Heebner
Dr. Dennis R. Horn
Dr. Barry M. Horowitz
Maj. Gen. Ralph H. Jacobson, USAF (Ret.)
Dr. Anita K. Jones
Dr. Paul G. Kaminski
Adm. Isaac C. Kidd Jr., USN (Ret.)
Mr. Sol Love
Dr. A. Louis Medin
Dr. John M. Palms
Dr. Victor H. Reis
Prof. Thomas F. Rosenbaum
Lt. Gen. Philip D. Shutler, USMC (Ret.)
Gen. Donn A. Starry, USA (Ret.)
Gen. John W. Vessey Jr., USA (Ret.)

[Slide 13] Now I am going to discuss the recommendations of the panel and it is a very well know panel with a set of experience that covers all of these topics in depth. I will highlight some instructional things that were said by individuals of the panel since I thought they were useful. Norm Augustine made the point that we should think of simulation not as something that confirms or rejects a hypothesis, but as something that is a mind extender; it makes us think about an area of concern that we would not have focused our attention on otherwise. As a result, simulation leads us into an area of evaluation that may turn out to be crucial that we might not have otherwise dealt with. Phil Shutler made the point in commenting on the notion of centrally managing the reuse and distribution of simulation, that it ignores a very important point that the designers of the simulation themselves carry all the knowledge about what went into it, what can be extrapolated, what cannot. Transferring the models without the designers would be a error because many of these codes are so large that it would impossible for another organization to understand all that went in. Jack Vessy, when looking at our recommendations, decided to quote George Orwell who said, "there are times when the first duty of responsible people is to restate the obvious" and that is how we view these recommendations, as a statement of the obvious.

An Evaluation Framework

- Consists of:
  - Measures of effectiveness
  - Environment and threat definitions
  - System assumptions
  - Role of testing
  - Planned M/S for evaluation
  - Establishing M/S credibility

Recommendations

[Slide 14] I need one chart prior to presenting recommendations to provide a definition of what we call an evaluation framework. Imagine setting up for each program, at the start, a set of measures of effectiveness, assumptions (like those threats and those environments), and how one plans to use testing and simulation as augmentations to each other over the life of this program. Now there is no doubt that everyone of these will be wrong the first time you do it, but also imagine that you were willing, in addition, to setting this up, to continue to refine it as you gain knowledge through the development planning process. We will call that set of information an evaluation framework.

M/S In Evaluation

Finding:

- Need greater emphasis on operational evaluation during the development process, including the OT&E community
- M/S can assist in defining an evaluation framework for a program

Recommendation:

- Involve the DOT&E/OT&E community to:
  - Participate early in the requirements process
  - Translate the operational requirements to an operational evaluation framework
  - Document the evaluation framework that identifies the roles of M/S and testing for evaluation objectives at each milestone
- CAUTION — The development community must take the lead in these efforts since they will implement the framework. The OT&E community must maintain their independence through the management chain

[Slide 15] Our first recommendation says that , as I stated right at the beginning, we think that we really need to emphasis the learning role for operational testing during development, and that we would like to see this occur and be mechanised through the setting up of these evaluation frameworks at the start of the programs. This will also provide assurance that we will not run into the surprise of assumptions as we did in the DIVAD example, or changes in the measures of effectiveness as we did in the Aquila example. To do that we have to involve the independent operational test people right from the start. We cannot let them just come in at the end. This raises two cautions. The first caution is they will not remain independent if they get involved in the programs earlier than when they do now. Our view is that confusing aloofness and independence is a big error. The value of the independence is that they have knowledge to provide and a management chain that gives them the independence to provide it. To keep them aloof is to lose something, not to gain something. We recommend that that confusion not dispell this recommendation.

The second caution is how can these operational testers layout this evaluation framework; they are a small group and they do not have the wherewithall or knowledge to do that. We are not recommending that; we are saying that the development community should take the lead in laying out an evaluation framework, but the community of operational people should be involved in agreeing to and continuing to modify the framework as the development goes on.

M/S In Evaluation

Finding:

- **The foundation of the acquisition process, the operational requirement and its translation into system terms, can be improved**
- **M/S can help by providing sensitivity analyses which help to isolate requirement areas needing concentrated evaluation**

Recommendation:

- **Service acquisition executives ensure that M/S excursion analyses are applied systematically to help reach and maintain agreement on major aspects of requirements and system performance, and to focus, not replace, testing**

[Slide 16] The second recommendation we made has to do with the idea that we do not want simulation to prove or disprove things, but we want it to isolate high sensitivity areas. Sensitivities that could make our view of utility wrong, if we are on the wrong side of that sensitivity curve. We want to establish an important role for simulation in doing those sensitivity analyses, and this being the method of focusing us on those early operational tests. So that has to go on right at the start of programs, and we think that the DoD has to set up a system to in fact ask for those. We do not want to see fixed point simulation results; we want to see excursion analyses which will then be the basis for deciding whether or not early operational testing should be done; focus not replace testing.

M/S In Evaluation

Findings:

- **A development program, as embodied in a specification and contract, may become overly rigid, restricting the willingness to evaluate and incorporate changes as threat, technology and knowledge evolve**
- **M/S is a tool for continual evaluation of potential changes**

Recommendation:

- **The USD(A) establish policy and provide guidance to the acquisition community for systematically reevaluating system specifications using M/S and test results**

[Slide 17] The third recommendation has to do with the development period which earlier on we called rigid. Now simulation is an evaluation tool, as I stated earlier, and if we have a management process that attempts to keep contracts and specifications fixed, then there is no room for evaluation, because the purpose of evaluation is to determine whether those are correct or not. We believe that the current state of the acquisition process is such that we stifle evaluation. Program managers are not motivated to find reasons that their programs are not correct, they are interested in stabilizing things. If we have a culture that does not want do evaluation, we can have all the evaluation tools in the world and we will not do evaluation. So this recommendation states that we think that OSD has to do something to change that culture. Now that is a very hard recommendation; we do not have a specific one switch you turn that fixes this, but nonetheless we think that is crucial if we are going to get any value out of the tools and the capabilities of simulation. This raises another caution. A set of people will say that, we will end up in a mode where we are always changing everything and we will end up with nothing. We will lose management control. In response to that we want to make it clear that this recommendation is not saying we should give up configuration management on system developments, and that every time someone does an evaluation that we should change something. There should be two distinct processes: one that is evaluating and one that is changing and we would expect the rate of evaluation is much higher than the rate of change. However, if we do not do any evaluation, there will be no changing, and then we will end up with those surprises that we talked about earlier.

M/S In Evaluation

Finding:

- **Accounting for human performance early in systems acquisition improves system capability and enhances the ability of the test and evaluation process to predict operational performance**

Recommendations:

- **Service acquisition executives ensure that the development programs employ man-in-the-loop simulation beginning with requirements definition and mature the simulation along with the hardware throughout the acquisition process**
- **JCS/CINCS exploit technology capabilities in distributed computing and networking to simulate coordinated combined arms engagements with man-in-the-loop and evaluate results against live exercises**

[Slide 18] The fourth recommendation has to do with the human factors problem illustrated by SACDIN. This just simply points out that the tools are now available and the cost is sufficiently low that every program should have an activity to build mock-ups of man machine interfaces as soon as possible, and bring in the real users to better understand the utility of that design. That should be upgraded as the fidelity of the development proceeds. Now there is one set of systems where this is particularly difficult to do, and that has to do with big command and control systems where the users are generals or admirals and there may be many of them. It is not easy to bring them in to try things out, and so here we are pointing to a set of technology that is emerging for doing what is called distributed simulation; that is, putting right in somebody's office his part to play in a global simulation and having a common war game underneath that simulation that multiple players located in various places could participate in concurrently. Perhaps through that mechanism we can get higher level people in the defense department to play in those portions of the systems that they in fact will ultimately be the user for. We are seeing that technology used today for training, but we think that it can start to move into the area of development as well.

Credibility Of M/S

[Slide 19] Finally I want to make some recommendations with regard to this issue of credibility.

Questions Raised On Credibility

- **How do we know whether or not to trust a simulation's output?**

  ANSWER: Excursion analyses, confirming test data, documentation on results

- **Would we gain confidence by setting up a central accreditation process for M/S?**

  ANSWER: No

- **Would we gain efficiency by setting up a central process to manage the use of M/S?**

  ANSWER: No

[Slide 20] I would like to do that by answering the questions we were asked first. How do we know whether to trust the simulation, should we set up this central office to accredit simulations, and should we set up a management process to distribute and reuse simulations? Our answers are as follows: on trusting simulation, we should not be looking for single point answers. We should be doing these excursion analyses, these sensitivity analyses. When we find something that makes us nervous, we should run a test and that is the validation, not the simulation, the test. We should have professionally documented simulation results. Often all we see is the one chart that we did a simulation and this is the answer, instead of seeing that whole set of data that makes someone understand how this simulation was calibrated, to what degree was the extrapolation of old data into new questions made, so that we have an ability to bring in a third party who could evaluation whether things seem credible or not. Should we set up this central office? The answer from the earlier discussion is a clear "no"; there is no office we think that could do it; it is not feasible. Should we set up a distribution process for redistributing these simulations? Well if we cannot accredit them we certainly cannot have a very logical process for redistributing them, so we say no to that to. But thats not to say that we are negative about credibility. We do have some recommendations.

Credibility Recommendations

- Support refinement, maintenance and availability of models which are reused by funding executive agents through JCS and OSD (e.g., DNA-Nuclear, DIA-Threat)
- Selectively use independent panels of experts to help evaluate M/S results, when confidence is unsure and decision is key
- DAB documentation should include
  - M/S plan and methodology
  - M/S limits, assumptions, extrapolations, sensitivities, inputs and outputs
  - Results analysis and test validation
- Use more than one model for the same analysis so comparisons can be made

[Slide 21] There are certain models in the defense department that tend to be used and reused alot by an expert group, such as the Defense Nuclear Agency (DNA) models on nuclear effects or the DIA threat models. Now interestingly enough in the DNA case these are never validated, and in the DIA case they often lead to alot of problems, but nonetheless, they are the best we have. Everybody uses them and those types of models should have a budget line to reinforce their improvement and keep them current. We recommend that the JCS and OSD in fact do that, but direct the money directly to those groups that in fact both develop and use the simulations.

If a subject comes up where simulation seems to be a crucial part of it, we think that there is a validation that can be done via an independent panel. You cannot validate the simulation itself, but you can validate a whole bunch of things like, the people who design the simulation, do they seem to know what they are talking about; the extrapolation, that is you can ask specific questions, to do with the historical use of the simulation for its validation versus the extrapolation you are doing now. The input data: are we putting garbage in and how do you know you are not. Is this simulation doing a partial evaluation or a full evaluation, and whether the parts you are not evaluating are more important than the parts you are? Those are all questions that we think panels of experts can in fact get at in a fairly short period of time, to add at least that level of confidence to the use of simulation when needed. That should only be done in very special cases. On the professional documentation point I made earlier, today's DAB documentation has a place for, but does not call for the data that determines what is the basis for validating simulation, and the credibility factors and so we are saying that those things in fact should become part of that documentation. Finally as that over-the-horizon radar example pointed out, you can sometimes find strange results occur by comparing two different validated models on the same extrapolated problem, and so this is at least a method of finding out whether things seem to be reasonably stable or not. That is what we have to say about credibility.

Summary

- Avoid Operational Test and Evaluation Surprises by:
  - Doing more testing and evaluation during development, for operational learning value
  - Setting up an Evaluation Framework for M/S and test at the start of programs, involving the independent operational testers. Reset the framework over time.
- Todays acquisition process stifles evaluation during development. Need a more open attitude, and management processes to encourage operational evaluations during development that may result in specification changes.
- Don't need an independent simulation office. It would not add confidence.
  - Simulation focuses testing through sensitivity analysis
  - Testing validates simulation

[Slide 22] So the summary of our study is that we think that it is really important to avoid those operational test surprises both for reasons of cost and credibility. The way we think you do that is by doing more operational testing during development so that we learn about the problem areas while they are still fixable, rather than wait until the very end. The way we see setting up a process for doing that is laying out evaluation frameworks which try to predict how we will do evaluation, but then as the real program progresses, upgrading them so that they are consistent with the state of knowledge. Have the operational testers involved so that we do not run any gaps of change that are unanticipated. The second point we make is the acquisition process stifles evaluation, and unless we have a more open attitude to doing evaluation it does not matter what our schemes are, we will not do it. So something has to be done at the upper management levels to change that. Finally we do not think you should set up independent simulation offices to accredit or manage the use or distribution of simulations. They will not add confidence, they would add confusion. We think that the way to think about this issue of confidence is that simulation focuses testing into those areas where we do not have confidence and helps us recognize those places where sensitivity is sufficiently questionable that it is worth testing. That way we get our confidence; through testing married with simulation.

BLANK PAGE