

Measures

RPG Special Topic

4 August 2004

Table of Contents

<u>Introduction</u>	1
<u>M&S Requirements and Measures</u>	1
What are Measures?	1
How are Measures “Measured”?	3
How are Measures Identified?	3
<u>The Dendritic Analysis Structure</u>	5
Overview	5
Implementing the Dendritic Approach	7
Establishing Measures and Acceptability Criteria	8
Comparing Simulations Results to Criteria	8
<u>Conclusion</u>	8
<u>References</u>	8
External Links and This Document	9
RPG References in This Document	9

This document corresponds to the web version of the VV&A RPG Special Topic of the same name and date. It has been modified to make it suitable for printing.

Introduction

This document discusses the use of measures in model and simulation (M&S) verification, validation, and accreditation (VV&A). Measures are extensively used in analysis and testing communities to assess Cost Effectiveness/Cost Benefit (CE/CB) issues in social sciences, medicine, labor/personnel, etc. The concepts described herein are consistent with the methodologies used in these communities.

There are two kinds of measures associated with VV&A:

- measures, or metrics, that describe the quality of the M&S product and (development) process and are used to provide information to facilitate their improvement
- functional measures that are used to assess some aspect of the subject represented in the simulation,¹ such as measures of performance (MOPs), measures of effectiveness (MOEs), and measures of merit (MOMs)

This document is concerned only with the second kind.

Measures are normally developed from an analysis of the M&S requirements and are based on the results of the operational risk assessment. Defining measures is a critical step in accreditation planning because they identify what is essential to the accreditation assessment and the subsequent accreditation decision. Clearly defined measures are critical to V&V planning because they identify the types and forms of information that need to be collected for the accreditation assessment, which, in turn, determine what V&V tasks are to be performed, what techniques should be used, and what resources will be required. Properly defined measures can also assist the simulation development or modification effort in assessing cost-benefit tradeoffs (i.e., balancing the desired simulation capabilities (maximum) against the risk (minimum) of using the simulation).

M&S Requirements and Measures

What are Measures?

Measures are derived from the M&S requirements of the intended use. To be useful in validation and accreditation, measures should be defined in terms that specify how they can be supported by data from V&V and accreditation activities.

¹ Throughout this document the term "simulation" will be used as a general term to indicate model, simulation, or federation.

There are three commonly used types of measures:

- **Measures of Merit (MOMs)** – measures that relate the effects of a concept or system to the mission that the concept or system supports

MOMs measure concept or system capabilities in terms of the effects of these capabilities on the overall mission of which the concept or system is a part. They measure mission attributes that define the overall objectives that the simulation must be capable of accomplishing.

Example:

An attribute for an aircraft flight simulator is realism. Measures for aircraft flight simulation realism include: visual fidelity, control feel, switch placement and functionality, etc.

If a MOM is defined in terms that do not permit direct evaluation (i.e., do not involve specific acceptability criteria), it should be supported by one or more MOEs or MOPs that can be evaluated. In some communities, MOMs are also known as mission level measures (MLMs).

- **Measures of Effectiveness (MOEs)** -- measures used to assess a system's effectiveness in the accomplishment of a task

MOEs measure simulation capabilities in terms of task accomplishment or simulated system attributes. Most simulations involve capabilities that can be related directly to operational capabilities in terms of engagement or battle outcome, e.g., Force Exchange Ratio, Hard Target Kills per Aircraft Sortie. For those that cannot be directly related to operational capabilities, the developed measures should be expressed in terms of concept or system attributes.

Example:

Measures for a Command and Control (C2) simulation may be expressed in terms of capacity, consistency, timeliness, accessibility, completeness, accuracy, transportability, and/or security. The measures selected for use should be developed to levels of specificity such that the C2 system can be assessed for its fitness for purpose.

MOE evaluation criteria (acceptability criteria) should be quantitative if at all possible.

- **Measures of Performance (MOPs)** -- measures used to gauge system or system component capabilities or characteristics

MOPs are quantitative or qualitative measures of simulation capabilities and characteristics. They are based on simulation capabilities and characteristics that are defined by the requirements of the intended application or that meet User-defined system performance requirements. Quantitative MOPs are used when it is difficult to directly assess an MOE or when quantitative criteria need to

be established. Qualitative MOPs are categorical measures of performance that refer to the presence or absence of specified characteristics. Quantitative MOPs can frequently be related to a numerical scale, such as a communication system simulation throughput. Subjective measurement techniques are generally used to address qualitative MOPs.

How are Measures “Measured”?

The act of determining, evaluating, or assessing each measure (e.g., MOP, MOE, MOM) results in a value that indicates how well the simulation is able to address the associated M&S requirement. Associated with each measure is a ***criterion*** that shows how well the measure needs to be addressed by the simulation in order to be acceptable for the intended use. These criteria are typically called ***acceptability criteria*** because they define the minimum level of performance, degree of effectiveness, level of success, etc. that the simulation needs to achieve to be acceptable to the User. Comparing measure results to their associated criteria determines if the simulation is acceptable. Acceptability criteria are normally defined by the User or Accreditation Agent for use during the accreditation assessment.

Acceptability criteria (or an appropriate subset) are also used in the V&V effort. During the validation process, the measured result values are compared with the acceptability criteria to assess the fitness of the simulation for the intended use. In cases where acceptability criteria have not been defined by the User or Accreditation Agent, the V&V Agent defines measures and criteria (frequently called *validation criteria*) to be used during validation. To ensure a thorough evaluation of a simulation’s credibility, acceptability criteria (or validation criteria) should be associated with every measure derived for the simulation.

Criteria are frequently expressed as thresholds – the minimum capability needed for the simulation to be fit for the intended purpose as defined and agreed to by the User. Criteria should be unambiguous and assessable, whether stated qualitatively or quantitatively. Often, these criteria are based on a predetermined standard or referent.².

How are Measures Identified?

For each intended use of a simulation, well-defined mission objectives and requirements should be articulated by the User. The VV&A team then develops measures and criteria for each simulation requirement or objective. The measures provide a way of relating the results of V&V and accreditation activities to the simulation’s ability to solve the problem or resolve the objectives (meet requirements). The progression from mission objectives to M&S requirements to measures and criteria is derived by a series of analysis steps:

² See the special topic on Fidelity for additional information on referents.

1) Obtain an accurate and complete statement of the problem, objectives and requirements from the User

An accurate and complete statement of the problem objectives and mission requirements is necessary in order to fully develop criteria by which the fitness of the simulation can be evaluated.

2) Define and refine (and verify) the M&S requirements

Once the problem objectives and requirements are well understood, the Accreditation Agent should assist the User in defining and refining the M&S requirements. The M&S requirements should be verified to ensure they completely and unambiguously elucidate the capabilities needed by the simulation to address the intended use.³

3) Conduct a risk assessment

The risk assessment provides a basis for planning for the VV&A effort by identifying the critical M&S requirements. The risk assessment should cover both operational and developmental risk, although greater emphasis should normally be placed on operational risk.⁴

4) Identify accreditation information needs

Once the M&S requirements have been defined and verified and risks have been assessed, the Accreditation Agent can identify and prioritize the information needed for the accreditation assessment.

5) Define V&V objectives to address the accreditation information needs

The V&V Agent uses the *accreditation information needs* to identify and prioritize the objectives of the V&V effort.

6) Decompose the V&V objectives until each is clearly defined in terms of specific measures, criteria, and tasks

V&V objectives should be defined in terms of the measures and specific tasks to be performed to address each measure. Objectives may need to be decomposed to attain this level of specificity. The ideal situation is when one V&V objective is defined by one measure that can be obtained performing one task. This situation seldom occurs. Generally, it will take many measures to address each objective, and multiple tasks to address each measure. However, where possible the principle of simplicity and direct traceability of objectives to sub-objectives, to measures, to tasks should be used as a guide in the full development of objectives.

³ See the special topic on Requirements for additional information.

⁴ See the special topic on Risk and Its Impact on VV&A for additional information.

Failure to solidify problem objectives and requirements, and to assess risks at the very start of planning for the VV&A effort will impact all subsequent tasks and activities undertaken by the V&V and Accreditation Agents.

These steps are more than fundamental, they are essential. The accreditation effort cannot attain meaningful results if problem objectives and M&S requirements are subject to continual change or revision. This is not to say that requirements and objectives cannot be changed at all, but rather that changes should occur only when absolutely necessary. In addition, M&S requirements should be reverified, and risks reassessed to address each change. The User is the final approval authority for any changes to M&S requirements.

The Dendritic Analysis Structure

Overview

Decomposition of missions into functional processes and finally into system attributes is a common analytical approach, sometimes referred to as a **dendritic** approach due to the dendrite or tree-like form of the resulting analytic structure. Unfortunately, this approach usually results in a complex analytical structure that defines many more compositional elements than a reasonable V&V effort could measure. The problem is simplified by considering only the critical elements of the mission or objective function. Inherently, this means that the elements of the problem are not equally weighted and the analyst (e.g., Accreditation Agent, User, V&V Agent) should conduct sensitivity analyses, early trade-off studies, or apply expert judgment to select (or deselect) the critical factors that will be measured and consequently used to drive the V&V effort and accreditation assessment. However, completeness is important. An incomplete dendritic structure may not yield enough information to adequately address the objectives. Worse yet, an incomplete structure may result in false conclusions drawn from incomplete information.

The dendritic process is a structured process that permits identification of all of the elements or actions necessary to address or resolve simulation objectives. The process provides the underlying structure for the VV&A analysis approach. The dendritic structure becomes the basis of other planning tasks, i.e., selection of V&V events and techniques and identification of resources needed to conduct the evaluation.⁵

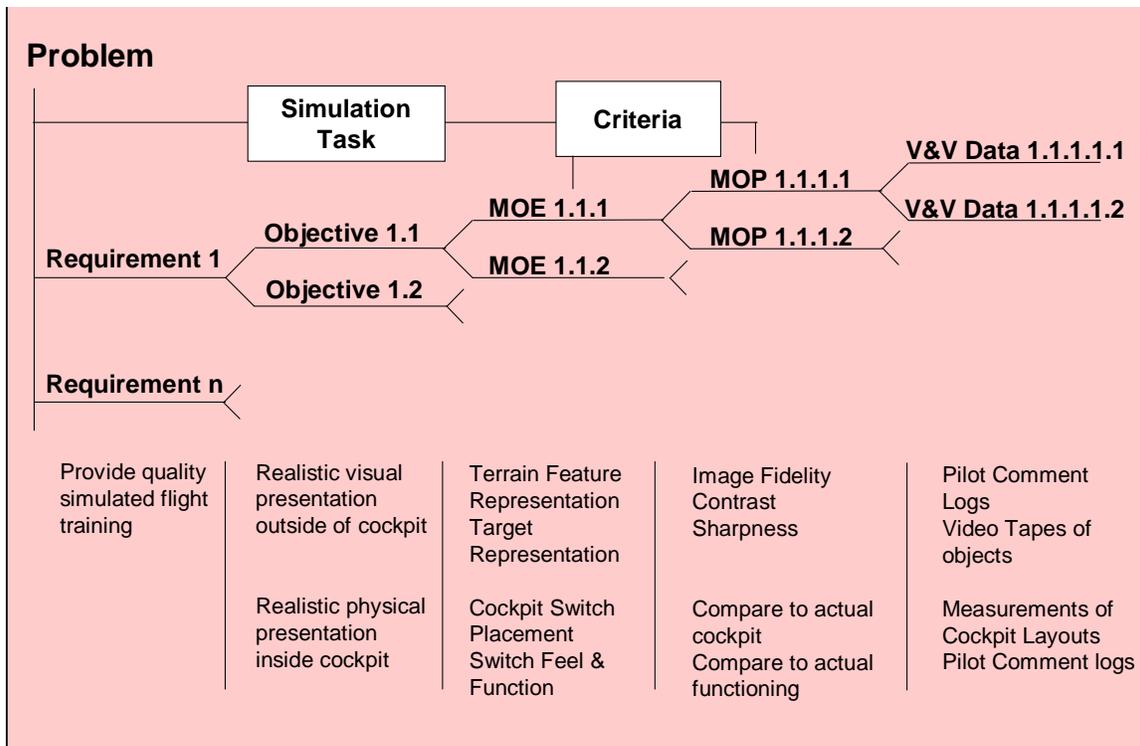
A key component to the process is linkage: all data collected using the various VV&A activities is needed and has a place in satisfying the objectives. This linkage assures that collection of only needed data is planned for and that resources will not be wasted by performing unnecessary activities or collecting unnecessary data.

⁵ For additional information on V&V and accreditation planning, see the core documents on the roles of the V&V Agent and Accreditation Agent, respectively.

Time and resources may preclude addressing every aspect of the concept or problem. Therefore, the approach should include sensitivity analysis and risk assessments of the key issues to identify those specific areas that show the greatest potential for impact on the accreditation decision. In addition, expert judgment is a practical necessity. The judgment of subject matter experts (SMEs) may be useful in developing or confirming the completeness of an analytical structure. This may allow the scope of early trade-off studies and sensitivity analyses to be minimized or focused on areas of greatest uncertainty. Judgment should be recognized as fundamental to the analytical process, and the rationale and impacts of analytical judgments should be consciously addressed.

One approach for defining simulation objectives is to establish a hierarchy in which the objectives (i.e., requirements) are at the highest level. These, in turn, are supported by sub-objectives and then measures. Many examples from Department of Defense (DoD) publications show a structure that develops objectives from the problem that the simulation is to solve, decomposes the objectives to a level that they can be addressed by measurable quantities, develops those measures, defines pass-fail criteria, and identifies supporting empirical data for comparison. The discipline of using this analytical hierarchy (dendritic) type of architecture may appear to thwart creativity, but to the contrary, the structure it contributes to the VV&A effort should help to illuminate areas that are ambiguous or uncertain and where creative thought is needed.

The thrust of this breakdown is to subdivide simulation objectives into more explicit measures or sub-measures that can be assessed against information collected during VV&A events. Measures may be formed at different levels depending on the objective. That is, one objective may have a sub-objective that leads immediately to quantifiable measures, while another objective may require additional subdivisions before measures are evident. The following figure provides an example set of levels for a structured approach to developing measures for a flight simulator.



Example set of Measures for a Flight Simulator

Implementing the Dendritic Approach

To implement this approach the analyst breaks the objective down into smaller and smaller segments by asking, "what do I need to know to answer this question?" Each level of the structure is given a title, i.e., requirement, objective, measure, or empirical data. However, there is no set number of levels.

This structured approach fosters discipline, accountability, and visibility. The resulting analytic structure preserves the organization and logic of why specific empirical data are required and, therefore, serves as an audit trail. The analyst's decisions in establishing each level help to focus the planning in a specific direction. The V&V Agent and Accreditation Agent should understand and consider how these decisions will impact the VV&A effort. The analytic structure will identify the need for collection of specific data during V&V and testing activities and for analyses to be performed in a particular manner. Thus, each decision made in building the analytic structure may have major resource considerations.

Performing this structured analytic process may appear to be simple and easily accomplished because it seems to imply "just continue to analyze and breakdown the problem into measurable parts." In fact, construction is often inductive, not deductive, and requires expert knowledge and often-substantial creativity.

Establishing Measures and Acceptability Criteria

Measures and acceptability criteria are critical to both the validation and accreditation efforts: They set the “pass/fail” data points for the entire VV&A effort. The Accreditation Agent (or the V&V Agent) analyzes the requirements, first deriving measures and then associating criteria with each measure. Other measures and criteria may be derived by SMEs. Criteria values may be collected from field test data, results of validation experiments, or predictions made by qualified SMEs (collectively called the referent). Criteria should be quantitative whenever practical, but may be supplemented by qualitative values provided by the User and SMEs. Criteria should reflect the overall needs of the intended application. Criteria that are set higher than necessary increase the cost of developing and validating the simulation; criteria that are set lower than necessary increase the risk that the results may not be credible. Ensuring that the values selected, quantitative and qualitative, are appropriate for the M&S requirements is a primary consideration of the V&V effort. Once the measures and criteria have been determined, they are submitted to the User for approval.

Comparing Simulation Results to Criteria

The bottom line in simulation accreditation is, “Does the simulation produce a credible result?” The principal source of information used to answer this question is results validation. The specification and use of measures and criteria comprise an important facet of this validation activity. In results validation, simulation results are provided in terms of the measures and these values are compared to the values of the associated acceptability (or validation) criteria. Measures should be devised that actively support the assessment of confidence for relevant simulation responses for the intended application. In addition, measures should allow for quantification of both errors and uncertainty in the comparison of simulation results to the referent.

Conclusion

Measures and associated criteria properly developed to address the M&S requirements of the intended application can provide a firm analysis basis from which the accreditation assessment can be made. A structured approach, such as the dendritic analysis structure, provides a rigorous methodology for developing measures and the criteria used with them.

References

DoD 5000.59-M, Department of Defense Modeling and Simulation Glossary, January 1998.

- Joint Test and Evaluation Program, Joint Feasibility Study (JFS) Handbook, Office of the UnderSecretary of Defense for Acquisition, Technology, and Logistics/Director for Test, Systems Engineering and Evaluation (OUSD(AT&L)/DTSE&E), 1996.
- Joint Crisis Action Test and Evaluation Program, “Program Test Design”, UnderSecretary of Defense for Acquisition and Technology/Director for Test and Evaluation (USD(A&T) DT&E), JCATE 93-1, 12 October 1993.
- Air Force Test and Evaluation Center (AFOTEC) Manual 99-101, “Test and Evaluation Business Management Process Manual”, 20 January 2000.
- Commander, Operational Test and Evaluation Force Instruction (COMOPTEVFORINST) 3960.1H, “Operational Test Director’s Guide”, 13 December 1995.
- U.S. Army Test and Evaluation Command (ATEC) Regulation 73-1, “System Test and Evaluation Policy”, 10 December 1999.
- Air Force Material Command (AFMC) “Detailed Test Plan Process Action Team Final Report”, August 1993.
- Invited Paper for Foundations in Verification and Validation in the 21st Century Workshop, “Verification, Validation, and Predictive Capability in Computational Engineering and Physics”, William Oberkampf, Timothy Trucano, and Charles Hirsch, presented 22 October 2002.
- Burge, Albert R., “Test and Evaluation Based on Metrics, Measures, Thresholds and Indicators,” *ITEA Journal of Test and Evaluation*, Vol. 17, No.3, 1996.

External References in This Document

University of California, School of Engineering, Information Sciences Institute (ISI)
website: <http://www.isi.edu/expect/measures-of-merit/node1.html>

RPG References in This Document

- select menu: *RPG Core Documents*, select item: “Role of the Accreditation Agent in the VV&A of Legacy Simulations”
- select menu: *RPG Core Documents*, select item: : “Role of the Accreditation Agent in the VV&A of New Simulations”
- select menu: *RPG Core Documents*, select item: : “Role of the V&V Agent in the VV&A of Legacy Simulations”
- select menu: *RPG Core Documents*, select item: “Role of the V&V Agent in the VV&A of New Simulations”
- select menu: *RPG Special Topics*, select item: “Fidelity”
- select menu: *RPG Special Topics*, select item: “Requirements”

select menu: RPG Special Topics, select item: "Risk and Its Impact on VV&A"

The appearance of hyperlinks does not constitute endorsement by the DoD, DMSO, the administrators of this web site, or the information, products or services contained therein. For other than authorized activities such as military exchanges and Morale, Welfare and Recreation sites, the DoD does not exercise any editorial control over the information you may find at these locations. Such links are provided consistent with the stated purpose of this DMSO web site.

§ § § § § § §