

VERIFICATION, VALIDATION, AND ACCREDITATION

Osman Balci

Department of Computer Science
660 McBryde Hall
Virginia Tech
Blacksburg, Virginia 24061, U.S.A.

ABSTRACT

This paper presents guidelines for conducting verification, validation, and accreditation (VV&A) of M&S applications. Fifteen guiding principles are introduced to help the researchers, practitioners and managers better comprehend what VV&A is all about. The VV&A activities are described in two M&S life cycles. Applicability of 77 V&V techniques is shown for the major stages of the two M&S life cycles. A methodology for accreditation of M&S applications is briefly introduced.

1 INTRODUCTION

Assuring total quality in a modeling and simulation (M&S) effort involves the measurement and assessment of a variety of quality characteristics such as accuracy, execution efficiency, maintainability, portability, reusability, and usability (human-computer interface). This paper is concerned only with the *accuracy* quality characteristic. Verification, validation, testing, accreditation, certification and credibility assessment activities primarily deal with the measurement and assessment of *accuracy* of models and simulations (M&S).

Model Verification is substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. Model verification deals with building the model *right*. The accuracy of transforming a problem formulation into a model specification or the accuracy of converting a model representation from a micro flowchart form into an executable computer program is evaluated in model verification.

Model Validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the M&S objectives. Model validation deals with building the *right* model.

An activity of accuracy assessment can be labeled as verification or validation based on an answer to the following question: In assessing the accuracy, is the model behavior compared with respect to the corresponding

system behavior through mental or computer execution? If the answer is "yes" then model validation is conducted; otherwise, it implies that the transformational accuracy is judged implying model verification.

Model Testing is ascertaining whether inaccuracies or errors exist in the model. In model testing, the model is subjected to test data or test cases to determine if it functions properly. "Test failed" implies the failure of the model, not the test. A test is devised and testing is conducted to perform either validation or verification or both. Some tests are devised to evaluate the behavioral accuracy (i.e., validity) of the model, and some tests are intended to judge the accuracy of model transformation from one form into another (verification). Sometimes, the whole process is called *model VV&T*.

Accreditation is "the official certification that a model or simulation is acceptable for use for a specific purpose." (DoD Directive 5000.59 <http://triton.dmsomil/docslib/mspolicy/directive.html>).

The purpose of this paper is to present guidelines for conducting VV&A. Section 2 presents VV&A principles. Section 3 describes the VV&A activities in two M&S life cycles. Applicability of 77 V&V techniques is tabulated in Section 4. Section 5 introduces a methodology for accreditation of M&S applications. Concluding remarks are given in Section 6.

2 VV&A PRINCIPLES

According to the Webster's dictionary, a principle is defined as "1. an accepted or professed rule of action or conduct. 2. a fundamental, primary, or general law or truth from which others are derived. 3. a fundamental doctrine or tenet; a distinctive ruling opinion." All three definitions above apply to the way the term "principle" is used herein.

Principles are important to understand the foundations of VV&A. The principles help the researchers, practitioners and managers better comprehend what VV&A is all about. They serve to provide the underpin-

nings for the 77 V&V techniques presented in Section 4 Understanding and applying these principles is crucially important for the success of an M&S effort.

The fifteen principles presented herein are established based on the experience described in the published literature and the author's experience during his V&V research since 1978. The principles are listed in Table 1 in no particular order. For detailed descriptions of these principles, please see (Balci 1998; DoD 1996).

3 VV&A IN TWO M&S LIFE CYCLES

VV&A activities in the M&S life cycle are depicted in Figure 1 (DoD 1996, p. 3-18) and in Figure 2 (Balci 1998). Figure 1 uses the DoD terminology and Figure 2 uses the non-DoD terminology. The fundamental phases of the life cycle are present in both figures.

V&V is not a phase or step in the M&S life cycle, but a continuous activity throughout the entire life cycle as enunciated by Principle 1 in Table 1. The life cycle should not be interpreted as strictly sequential. The sequential representation of some arrows is intended to show the direction of

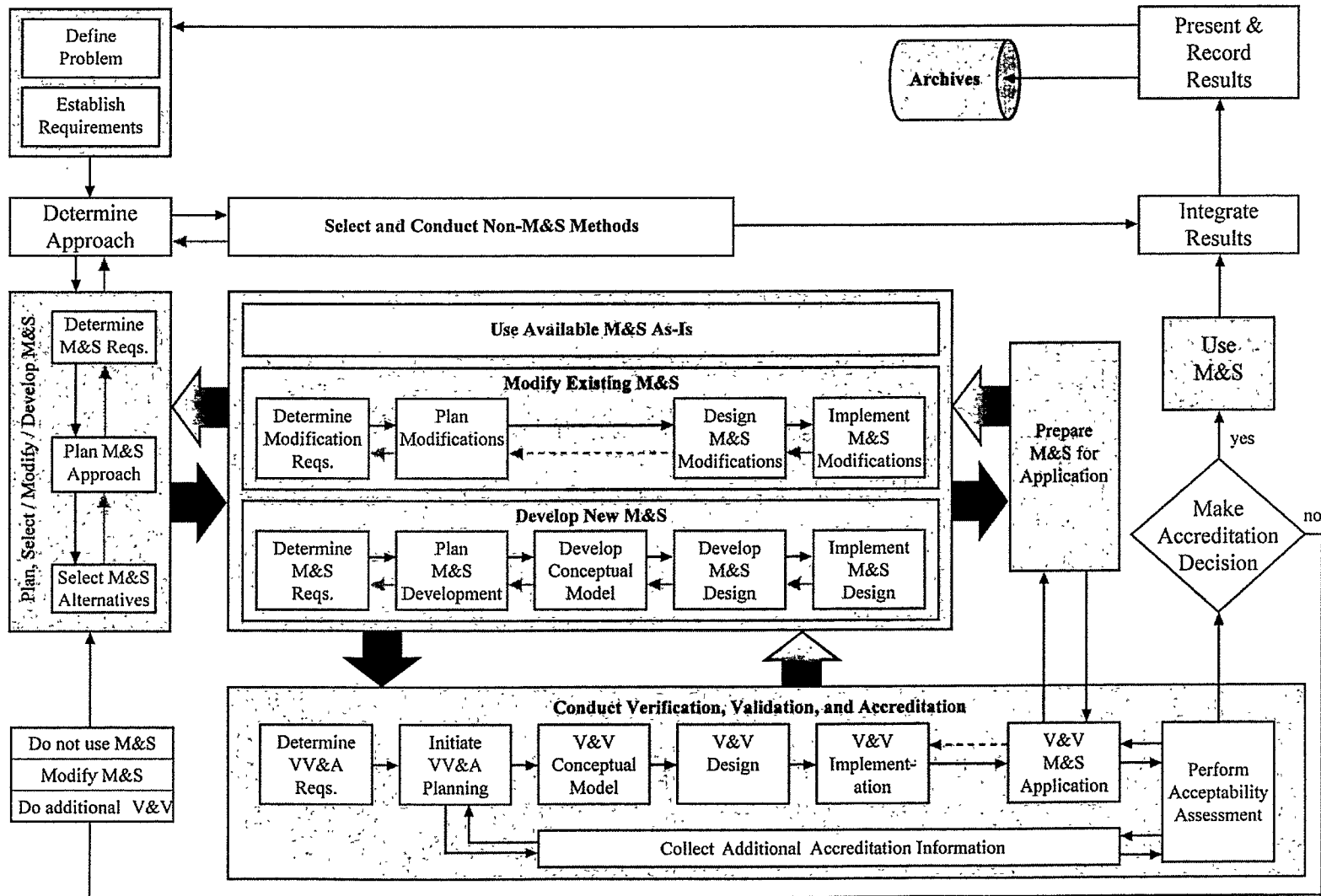
development throughout the life cycle. The life cycle is iterative in nature and reverse transitions are expected. Deficiencies identified by a VV&A activity may necessitate returning to an earlier process and starting all over again.

Conducting V&V for the first time in the life cycle when the M&S application is complete is analogous to the teacher who gives only a final examination (Hetzel 1984). No opportunity is provided throughout the semester to notify the student that he or she has serious deficiencies. Severe problems may go undetected until it is too late to do anything but fail the student. Frequent tests and homework throughout the semester are intended to inform the students about their deficiencies so that they can study more to improve their knowledge as the course progresses.

The situation in conducting V&V is exactly analogous. The VV&A activities throughout the entire M&S life cycle, shown in Figure 1 and Figure 2, are intended to reveal any quality deficiencies that might be present as the M&S progresses from the problem definition to the completion of the M&S application. This allows us to identify and rectify quality deficiencies during the life cycle phase in which they occur.

Table 1: Principles of VV&A

1	V&V must be conducted throughout the entire M&S life cycle.
2	The outcome of VV&A should not be considered as a binary variable where the model or simulation is absolutely correct or absolutely incorrect.
3	A simulation model is built with respect to the M&S objectives and its credibility is judged with respect to those objectives.
4	V&V requires independence to prevent developer's bias.
5	VV&A is difficult and requires creativity and insight.
6	Credibility can be claimed only for the prescribed conditions for which the model or simulation is verified, validated and accredited.
7	Complete simulation model testing is not possible.
8	VV&A must be planned and documented.
9	Type I, II and III errors must be prevented.
10	Errors should be detected as early as possible in the M&S life cycle.
11	Multiple response problem must be recognized and resolved properly.
12	Successfully testing each submodel (module) does not imply overall model credibility.
13	Double validation problem must be recognized and resolved properly.
14	Simulation model validity does not guarantee the credibility and acceptability of simulation results.
15	A well-formulated problem is essential to the acceptability and accreditation of M&S results.



Verification, Validation, and Accreditation

Figure 1: VV&A in the M&S Life Cycle (DoD 1996, p. 3-18)

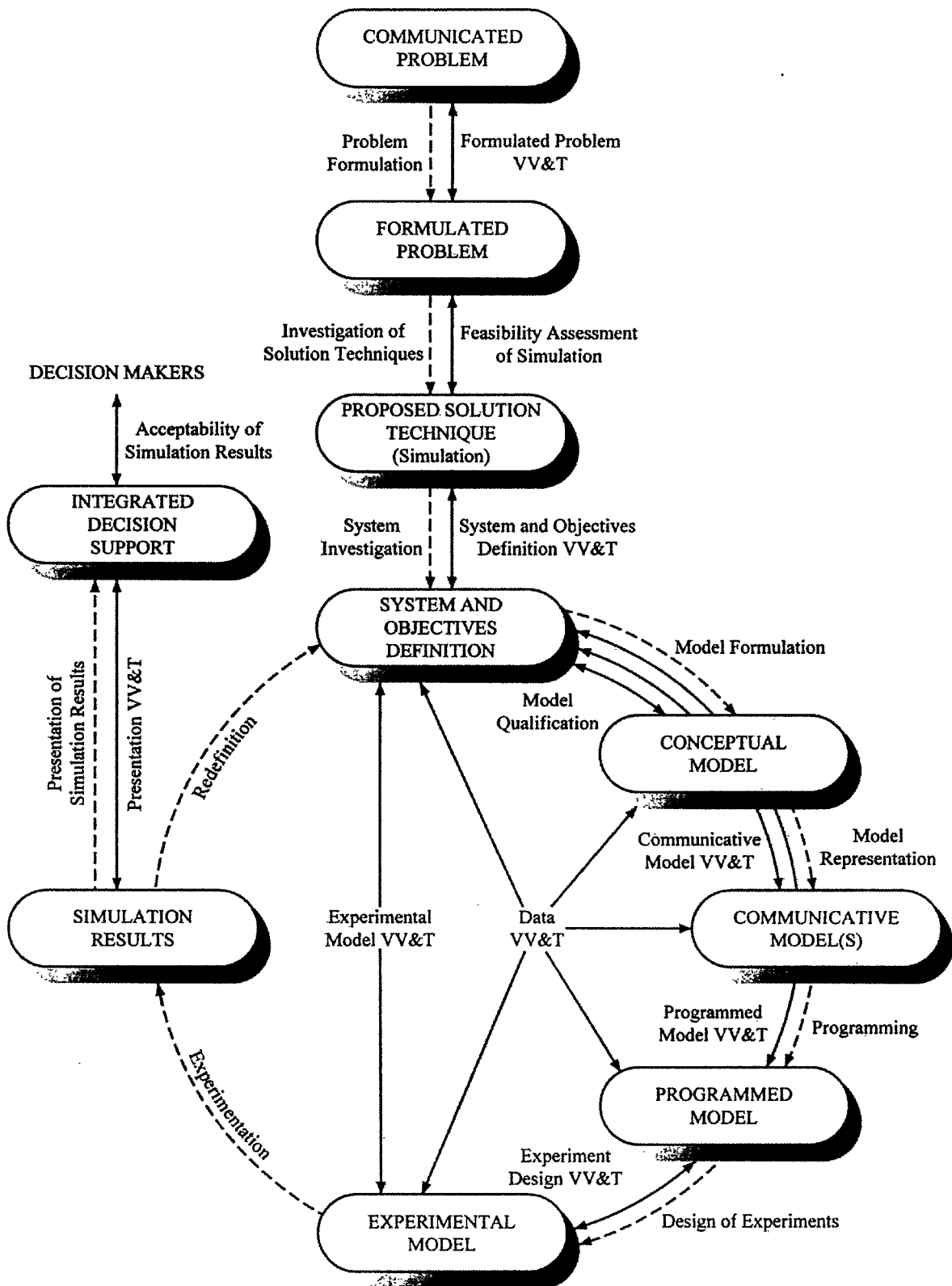


Figure 2: VV&A in Another M&S Life Cycle (Balci 1998)

As enunciated by Principle 10 in Table 1, errors should be detected as early as possible in the M&S life cycle. Delaying V&V to later stages in the life cycle increases the probability of committing errors.

4 APPLICABILITY OF THE V&V TECHNIQUES

Table 2 defines the column numbers of Table 3 for the M&S life cycles shown in Figures 1 and 2. Each column number represents a major stage, as defined in Table 2, in the M&S life cycle for V&V. It should be noted that Table 2 shows only the major stages and many other VV&A activities exist throughout each of the two life cycles.

Table 3 marks the V&V techniques that are applicable for each major stage of the M&S life cycle in Figures 1 and 2. The rows of Table 3 list the V&V techniques in alphabetical order. The column numbers correspond to the major stages in the life cycle.

The question of which of the applicable V&V techniques should be selected for a particular V&V activity in the life cycle should be answered by taking the following into consideration: (a) model type, (b) simulation type, (c) problem domain, and (d) M&S objectives.

Table 2: Index for Table 3

No	Figure 1	Figure 2
1	Problem Definition	Formulated Problem
2	M&S Approach	Feasibility Assessment of Simulation
3	M&S Requirements	System and Objectives Definition
4	Conceptual Model	Model Qualification
5	M&S Design	Communicative Model
6	M&S Implementation	Programmed Model
7	M&S Application	Experiment Design
8	M&S Acceptability Assessment	Data
9		Experimental Model
10		Presentation

Table 3: Applicability of the V&V Techniques Throughout the M&S Life Cycle

	Figure 1								Figure 2									
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
Acceptance Testing							•	•										•
Alpha Testing						•	•											•
Assertion Checking						•								•	•	•	•	
Audit	•	•	•	•	•	•			•	•	•	•	•	•	•	•	•	•
Authorization Testing						•	•	•						•				•
Beta Testing						•	•											•
Bottom-Up Testing						•								•				•
Boundary Value Testing						•								•				•
Branch Testing						•								•	•			•
Calling Structure Analysis				•	•	•								•	•			•
Cause-Effect Graphing	•		•	•	•	•			•					•	•			•
Comparison Testing		•	•	•	•	•	•											•
Concurrent Process Analysis						•	•							•	•			•
Condition Testing						•								•	•			•
Control Flow Analysis				•	•	•								•	•	•		•
Data Dependency Analysis	•			•	•	•								•	•	•	•	•
Data Flow Analysis				•	•	•								•	•	•	•	•

Table 3: Applicability of the V&V Techniques Throughout the M&S Life Cycle

	Figure 1								Figure 2									
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
Data Flow Testing						•								•	•	•	•	
Data Interface Testing						•	•									•	•	
Debugging						•								•	•		•	
Desk Checking	•	•	•	•	•	•			•	•	•	•	•	•	•	•	•	•
Documentation Checking	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Equivalence Partitioning Testing						•								•			•	
Execution Monitoring						•	•							•	•		•	
Execution Profiling						•	•							•	•		•	
Execution Tracing						•	•							•	•		•	
Extreme Input Testing						•								•			•	
Face Validation	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Fault/Failure Analysis						•	•						•	•	•	•	•	
Fault/Failure Insertion Testing						•	•							•	•	•	•	
Field Testing							•										•	
Functional Testing						•	•							•	•		•	
Graphical Comparisons						•	•							•	•		•	
Induction				•	•									•			•	
Inductive Assertions				•	•									•			•	
Inference				•	•									•			•	
Inspections	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Invalid Input Testing						•	•							•			•	
Lambda Calculus				•	•									•			•	
Logical Deduction				•	•									•			•	
Loop Testing						•								•	•		•	
Model Interface Analysis			•	•	•								•	•			•	
Model Interface Testing						•	•	•						•			•	
Object-Flow Testing						•	•	•						•			•	
Partition Testing					•	•								•	•		•	
Path Testing						•	•							•	•		•	
Performance Testing							•	•									•	
Predicate Calculus				•	•									•			•	
Predicate Transformation				•	•									•			•	
Predictive Validation						•	•	•									•	
Product Testing							•	•									•	
Proof of Correctness				•	•									•			•	

Table 3: Applicability of the V&V Techniques Throughout the M&S Life Cycle

	Figure 1								Figure 2									
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
Real-Time Input Testing						•	•	•						•			•	
Regression Testing						•	•							•			•	
Reviews	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Security Testing							•	•									•	
Self-Driven Input Testing						•	•	•						•			•	
Semantic Analysis					•	•								•	•		•	
Sensitivity Analysis						•	•	•						•	•		•	
Standards Testing							•	•						•			•	
State Transition Analysis				•	•	•							•	•	•		•	
Statement Testing						•								•	•		•	
Statistical Techniques						•	•	•								•	•	
Stress Testing						•	•	•						•			•	
Structural Analysis				•	•								•	•	•		•	
Submodel/Module Testing						•								•	•		•	
Symbolic Debugging						•								•	•		•	
Symbolic Evaluation					•								•	•	•		•	
Syntax Analysis						•								•	•		•	
Top-Down Testing						•								•	•		•	
Trace-Driven Input Testing						•	•	•						•			•	
Traceability Assessment				•	•	•	•						•	•			•	
Turing Test						•	•	•									•	
User Interface Analysis				•	•	•							•	•			•	
User Interface Testing						•	•	•						•			•	
Visualization/Animation						•	•	•						•	•	•	•	•
Walkthroughs	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

5 A METHODOLOGY FOR THE ACCREDITATION OF M&S APPLICATIONS

Accreditation requires (a) the measurement and evaluation of qualitative and quantitative elements of an M&S application, (b) expert knowledge, (c) independent evaluation, and (c) comprehensive assessment. We have developed a methodology satisfying these requirements. We have also developed a software tool that provides computer-aided assistance for the application of the methodology. Both the methodology and the software tool have been developed under U.S. Navy funding.

The multifaceted methodology employs a hierarchy of hundreds of indicators, analytic hierarchy process (AHP) for criticality weighting of indicators, fuzzy arithmetic, rule-based expert knowledge base, and an independent evaluation scenario using domain experts.

The software tool enables the creation of repositories of indicator hierarchies, criticality weighting of expert evaluators and indicators using different techniques, evaluation report generation in HTML, aggregation of crisp and fuzzy scores, and rule-based knowledge specification using an object-oriented scripting language.

6 CONCLUDING REMARKS

The life cycle application of V&V is extremely important for successful completion of complex and large-scale M&S efforts. This point must be clearly understood by the sponsor of the M&S effort and the organization conducting the M&S. The sponsor must furnish funds under the contractual agreement and require the contractor to apply V&V throughout the entire M&S life cycle.

Assessing credibility throughout the life cycle is an onerous task. Applying the V&V techniques throughout the life cycle is time consuming and costly. In practice, under time pressure to complete an M&S effort, the V&V and documentation are sacrificed first. Computer-aided assistance for credibility assessment is required to alleviate these problems. More research is needed to bring automation to the application of V&V techniques.

How much to test or when to stop testing depends on the M&S objectives. The testing should continue until sufficient confidence is achieved in credibility and acceptability of M&S results. The sufficiency of the confidence is dictated by the M&S objectives.

Establishing a simulation quality assurance (SQA) program within the organization conducting the M&S effort is extremely important for successful credibility assessment. The SQA management structure goes beyond V&V and is also responsible for assessing other model quality characteristics such as maintainability, reusability, and usability (human-computer interface). The management of the SQA program and the management of the simulation project must be independent of each other and neither should be able to overrule the other.

Subjectivity is, and will always be, part of the credibility assessment for a reasonably complex simulation study. The reason for subjectivity is two-fold: modeling is an art and credibility assessment is situation dependent.

REFERENCES

- Balci, O. 1998. Verification, validation, and testing. In *The Handbook of Simulation*, Chapter 10, J. Banks, Ed., John Wiley & Sons, New York, NY, to appear.
- Department of Defense. 1996. *Department of Defense Verification, Validation and Accreditation (VV&A) Recommended Practices Guide*, Defense Modeling and Simulation Office, Alexandria, VA, Nov. (Co-authored by: O. Balci, P. A. Glasow, P. Muessig, E. H. Page, J. Sikora, S. Solick, and S. Youngblood)
<http://triton.dmsso.mil/docslib/mspolicy/vva/rpg/>
- Hetzel, W. 1984. *The Complete Guide to Software Testing*, QED Information Sciences, Wellesley, MA.

AUTHOR BIOGRAPHY

OSMAN BALCI is an Associate Professor of Computer Science at Virginia Tech. He received B.S. and M.S. degrees from Bogazici University in 1975 and 1977, and M.S. and Ph.D. degrees from Syracuse University in 1978 and 1981. Dr. Balci is the Editor-in-Chief of two international journals: *Annals of Software Engineering* and *World Wide Web*; Verification, Validation, and Accreditation (VV&A) Area Editor of *ACM Transactions on Modeling and Computer Simulation* (TOMACS); Simulation and Modeling Category Editor of *ACM Computing Reviews*; and serves on five other editorial boards. He is a Director at Large of the Society for Computer Simulation (SCS). He will be a member of the WSC Board of Directors representing SCS in Jan. 1999. He has served as: Guest Editor of TOMACS for a special issue on "Simulation for Training: Foundations and Techniques"; Proceedings Editor of the 1990 Winter Simulation Conference; Program Chairman and Proceedings Editor of the SCS conference on Simulation Methodology and Validation, 1986-87; and Vice Chairman of ACM SIGSIM, 1985-87. He has published numerous technical articles and edited/co-edited four books. Dr. Balci has been a PI or Co-PI on research grants and contracts sponsored by the U.S. Navy with a total funding of \$1.3 million. His current research interests center on software engineering, visual simulation and modeling, and world wide web. Dr. Balci is a member of Alpha Pi Mu, Sigma Xi, Upsilon Pi Epsilon, ACM, IEEE CS, INFORMS, and SCS.