Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

VERIFICATION, VALIDATION, AND CERTIFICATION OF MODELING AND SIMULATION APPLICATIONS

Osman Balci

Department of Computer Science
660 McBryde Hall, MC 0106
Virginia Tech
Blacksburg, VA 24061, U.S.A.

ABSTRACT

Certifying that a large-scale complex modeling and simulation (M&S) application can be used for a set of specific purposes is an onerous task, which involves complex evaluation processes throughout the entire M&S development life cycle. The evaluation processes consist of verification and validation activities, quality assurance, assessment of qualitative and quantitative elements, assessments by subject matter experts, and integration of disparate measurements and assessments. Planning, managing, and conducting the evaluation processes require a disciplined life-cycle approach and should not be performed in an *ad hoc* manner. The purpose of this tutorial paper is to present structured evaluation processes throughout the entire M&S development life cycle. Engineers, analysts, and managers can execute the evaluation processes presented herein to be able to formulate a certification decision for a large-scale complex M&S application.

1 INTRODUCTION

A *model* is a representation or abstraction of something such as an entity, a system or an idea. *Simulation* is the act of experimenting with or exercising a model or a number of models under diverse objectives including acquisition, analysis, and training. For example, if the analysis objective is to predict the performance of a complex system design, we *experiment* with a model or a distributed set of models representing the system design. If the predicted performance is used in making an acquisition decision, the process is called *simulation-based acquisition.* If the training objective is to teach military commanders how to make decisions under a combat scenario, we *exercise* a model or a distributed set of models in an interactive manner by using the trainees as part of the simulation. We refer to a specific simulation created for a particular objective as a modeling and simulation (M&S) application.

Many types of M&S applications exist such as continuous, discrete-event, distributed, hardware-in-the-loop, software-in-the-loop, human-in-the-loop, Monte Carlo, and synthetic environments. Each M&S application type poses its own technical challenges for Verification, Validation, and Certification.

The terms Verification and Validation (V&V) are consistently defined for whatever entity they are applied to. Let *x* be that entity such as model, simulation, software, data, expert system, or a life-cycle artifact (product) such as requirements specification, conceptual model, design specification, or executable module. Then, V&V can be defined generically as follows:

- *x Verification* deals with the assessment of *transformational accuracy* of the *x* and addresses the question of "Are we creating the *x* right?"
- *x Validation* deals with the assessment of *behavioral or representational accuracy* of the *x* and addresses the question of "Are we creating the right *x*?"

For whatever entity to be subjected to V&V, substitute the entity name in place of *x* above, the definitions will hold.

Accreditation is defined in the Department of Defense (DoD) M&S community as "the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose" (DoDI 1996). On the other hand, the International Organization for Standardization (ISO) defines accreditation and certification as follows (Rae, Robert, and Hausen 1995):

- *Accreditation* is a "procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks."
- *Certification* is a "procedure by which a third party gives written assurance that a product, process or service conforms to specified characteristics."

The above ISO definitions conflict with the definitions commonly used by the DoD M&S community. To the best of our knowledge, all engineering disciplines, educational sector, and other areas use these terms as defined by ISO. Therefore, we use the ISO terminology in this paper.

2 PROPOSED PRACTICE FOR M&S ACCREDITATION AND CERTIFICATION

Similar to the manner accreditation and certification are carried out in engineering disciplines, educational sector, and other areas, we propose the comprehensive scheme shown in Figure 1 (Balci et al. 2002b; Balci and Saadi 2002).

We envision an accreditation authority at the national level. Example accreditation authorities include the United Kingdom Accreditation Service (http://www.ukas.com), Japan Accreditation Board for Conformity Assessment (http://www.jab.or.jp), and German Accreditation Council (http://www.dar.bam.de/indexe.html). We believe that the National Institute of Standards and Technology (NIST) can serve as the accreditation authority in the United States.

Under our proposed practice, those companies or organizations interested in serving as M&S certification agents apply to the accreditation authority. The accreditation authority examines the maturity of the applicant's standard certification processes and the qualifications of the key personnel who will execute the certification processes. Based on the examination results, the accreditation authority gives formal recognition that the applicant agent is competent to carry out the standard processes and provide certification which is unbiased, fair, cost effective, and consistent.

For example, many companies serve as ISO 9000 certification agents. These agents examine the processes of a company and certify that the company is compliant with the ISO 9000 standard. The ISO 9000 certification agents are accredited by an accreditation authority. A directory of ISO 9000 accreditation bodies is provided at (http://www.praxiom.com/accrediters.htm).

As the ISO definition indicates, certification must be conducted by a third party, where the first party refers to M&S application sponsor and the second party refers to M&S application developer. Of course, certification is meaningful when conducted in an independent manner. To achieve true independence, the IEEE Standard 1012 (IEEE 1998) requires technical, managerial, and financial independence as described below.

- *Technical Independence* implies that the certification agent determines, prioritizes, and schedules its own tasks and efforts.
- *Managerial Independence* implies that the certification agent reports to the M&S application sponsor independently of the developer organization.
- *Financial Independence* implies that the certification agent is allocated its own budget for the M&S certification and does not rely on the M&S development budget.

The accreditation authority establishes and publishes a set of criteria for accrediting M&S certification agents. The criteria should include:

- maturity of the applicant's standard certification processes,
- credentials of the key personnel who will execute the certification processes, and
- true independence of the agent.

The proposed practice is needed to provide the checks and balances required to minimize the M&S application sponsor's risks.

Certification is the independent award of a "Certificate", a "Seal of Approval" or a "Mark of Conformity" formally attesting that an M&S application fulfills specific quality criteria under a set of prescribed intended uses. The independent award is regarded by the M&S application

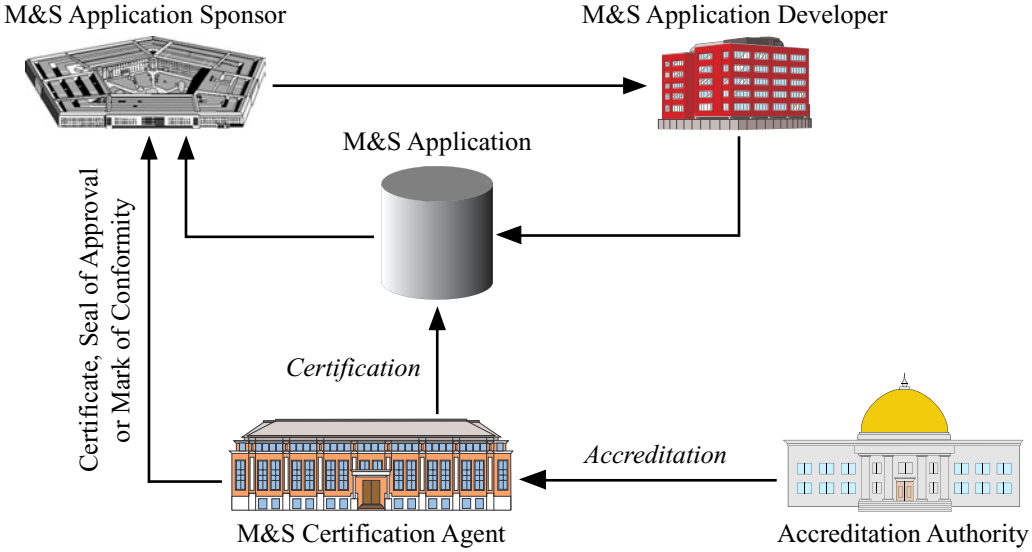

Figure 1: Proposed Practice for Accreditation and Certification

sponsor as providing some form of guarantee of quality and credibility. Based on the guarantee, the sponsor decides to use the M&S results in making key decisions. The consequences of wrongly awarding a "Certificate", a "Seal of Approval" or a "Mark of Conformity" may be catastrophic.

3 M&S LIFE CYCLE EVALUATION PROCESSES FOR CERTIFICATION

An M&S development life cycle describes the blueprint of development and specifies the work products to be created under the designated processes together with the integrated V&V activities. A life cycle is critically needed to modularize and structure the development process and provide valuable guidance for project management. A life cycle is also required to show the V&V activities as integrated within the development process based on the principle which dictates that V&V must go hand in hand with the development life cycle (Balci 1997).

An M&S development life cycle with evaluation processes for concurrent certification is presented in Figure 2. The life cycle consists of seven stages. A *stage* is defined to have an input work product (or artifact), an output work product (or artifact), and a process used to create the output product from the input product. A stage is conducted depending on the life cycle model employed. Many life cycle models exist including exploratory development, incremental development, prototyping, reuse-based development, spiral, and waterfall.

Although an evaluation process is named after the output work product of the corresponding stage, the evaluation process must assess more than the output work product. An *evaluation process* is defined to measure and assess a particular life cycle stage's (a) output work *product*, (b) *process* used in creating the output work product, and (c) *project* characteristics (i.e., people, documentation, planning, quality assurance, capability maturity).

The three Ps (Product, Process, Project) of software engineering are commonly referred to for software measurement and certification. Voas (1999) presents a software quality certification triangle, which includes the three Ps as Product, Process, and Personnel, and advocates that certification can be approached from any one of these aspects, but a combination of all three will provide the best balance.

Product quality is the degree to which the product possesses a desired set of characteristics. The first product quality characteristic "product accuracy" is assessed by evaluating product verity and validity. Product verity is evaluated by conducting product verification and product validity is evaluated by conducting product validation. We refer to product verification and product validation as simply V&V throughout the development life cycle.

Process quality is the degree to which the process possesses a desired set of characteristics. The set of desired characteristics depends on the process methodologies and techniques employed by the M&S application developer.

Project quality is the degree to which the project possesses a desired set of characteristics. Project quality is assessed by evaluating a variety of characteristics including configuration management, documentation quality, human resource management, personnel capability maturity, planning quality, and quality management.

Certification is a confidence building activity and can be best carried out if all three Ps are included. Concurrent certification, V&V, and quality assessment must be conducted in a manner integrated within the development life cycle as depicted in Figure 2.

The life cycle evaluation processes can be conducted under the guidance of the evaluation methodology presented by Balci (2001). The methodology enables the decomposition of a complex evaluation problem into a hierarchy of indicators in the form of an acyclic graph. The top-down hierarchical decomposition / modularization breaks the measurement complexity into small pieces corresponding to leaf indicators. Each leaf indicator is manageable in complexity and is directly assessable by way of testing, direct measurement, analysis, or examination. Only the leaf indicators need to be measured and evaluated.

The Evaluation Environment (EE) software system (Orca 2003) can be used to facilitate the application of the evaluation methodology (Balci et al. 2002a). EE is a Web-based client/server distributed software system structured based on the Java 2 Enterprise Edition (J2EE) industry-standard architecture. The 128-bit encrypted Secure Sockets Layer (SSL) technology is used to provide secure communication between the EE server and the EE user. EE enables geographically dispersed people to conduct complex evaluations in a *collaborative* manner.

3.1 Conceptual Model Evaluation Process

A simulation *Conceptual Model* is the model formulated in the mind of the modeler and specified in a variety of communicative forms intended for different users such as managers, analysts, and developers. Example communicative forms include animation, audio, chart, diagram, drawing, equation, graph, image, text, and video.

The *conceptual modeling* process involves the examination of the universe of discourse, problem formulation, system definition, identification of the Stakeholders and their needs, and specification of Intended Uses (Balci and Ormsby 2000) for which the M&S application will be built. Simulation conceptual models can be used (a) as a tool to control M&S requirements and content by the M&S manager, (b) to evaluate simulation concepts for thoroughness and errors, (c) as a foundation for M&S design, and (d) to verify and validate the M&S design (DMSO 2000).

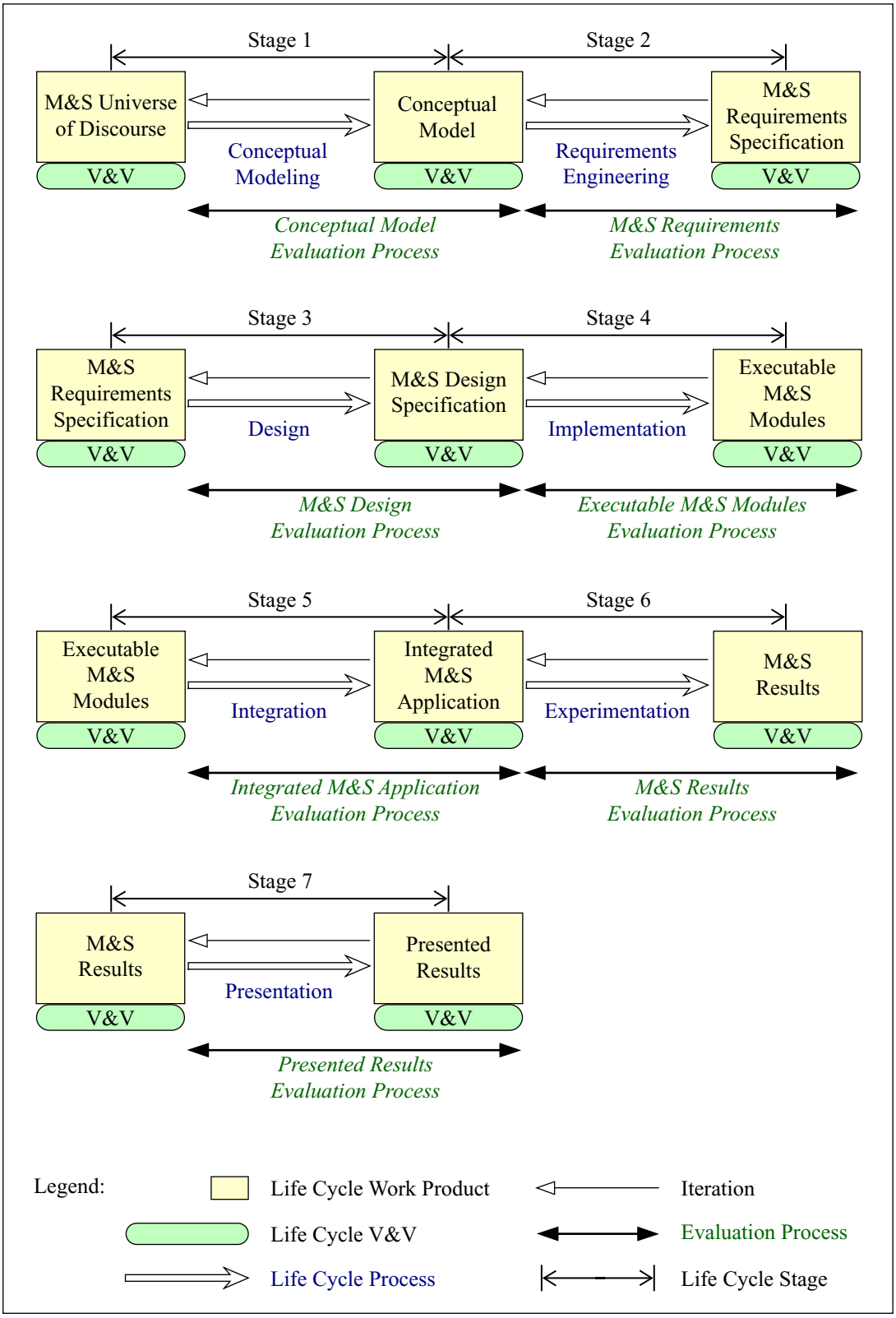

Figure 2: M&S Development Life Cycle Evaluation Processes for Certification

The *conceptual model evaluation process* assesses the credibility of the conceptual model created by conducting the conceptual modeling process. The process integrates the evaluations of: (a) conceptual model quality, (b) conceptual modeling process, and (c) M&S project characteristics related to life cycle stage 1.

3.2 M&S Requirements Evaluation Process

Properly identifying and specifying requirements is critically important. The *M&S Requirements Specification* document becomes the point of reference during the rest of the development life cycle. If the point of reference (i.e., requirements) is improperly formulated, all subsequent references made to the requirements may be erroneous.

Based on the Intended Uses specified in the Conceptual Model and the Statement of Work (SOW) for a particular M&S application release, *use cases* should be identified and the requirements should be specified for each use case. Use case-based requirements elicitation and specification is considered to be the industry best practice. The Rational Unified Process and the Unified Modeling Language (UML) technology (Rational 2003) advocate the identification and specification of functional requirements based on the use cases.

The *M&S requirements evaluation process* evaluates the credibility of the M&S requirements created by conducting the requirements engineering process based on the simulation conceptual model. The process integrates the evaluations of: (a) M&S requirements quality, (b) requirements engineering process, and (c) M&S project characteristics related to life cycle stage 2.

A life cycle for M&S Requirements Engineering is presented in Figure 3. The V&V activities shown in Figure 3 are carried out as part of the M&S requirements evaluation process.

M&S requirements can be evaluated in a collaborative manner as part of the overall M&S acceptability assessment by using an EE project as depicted in Figure 4.

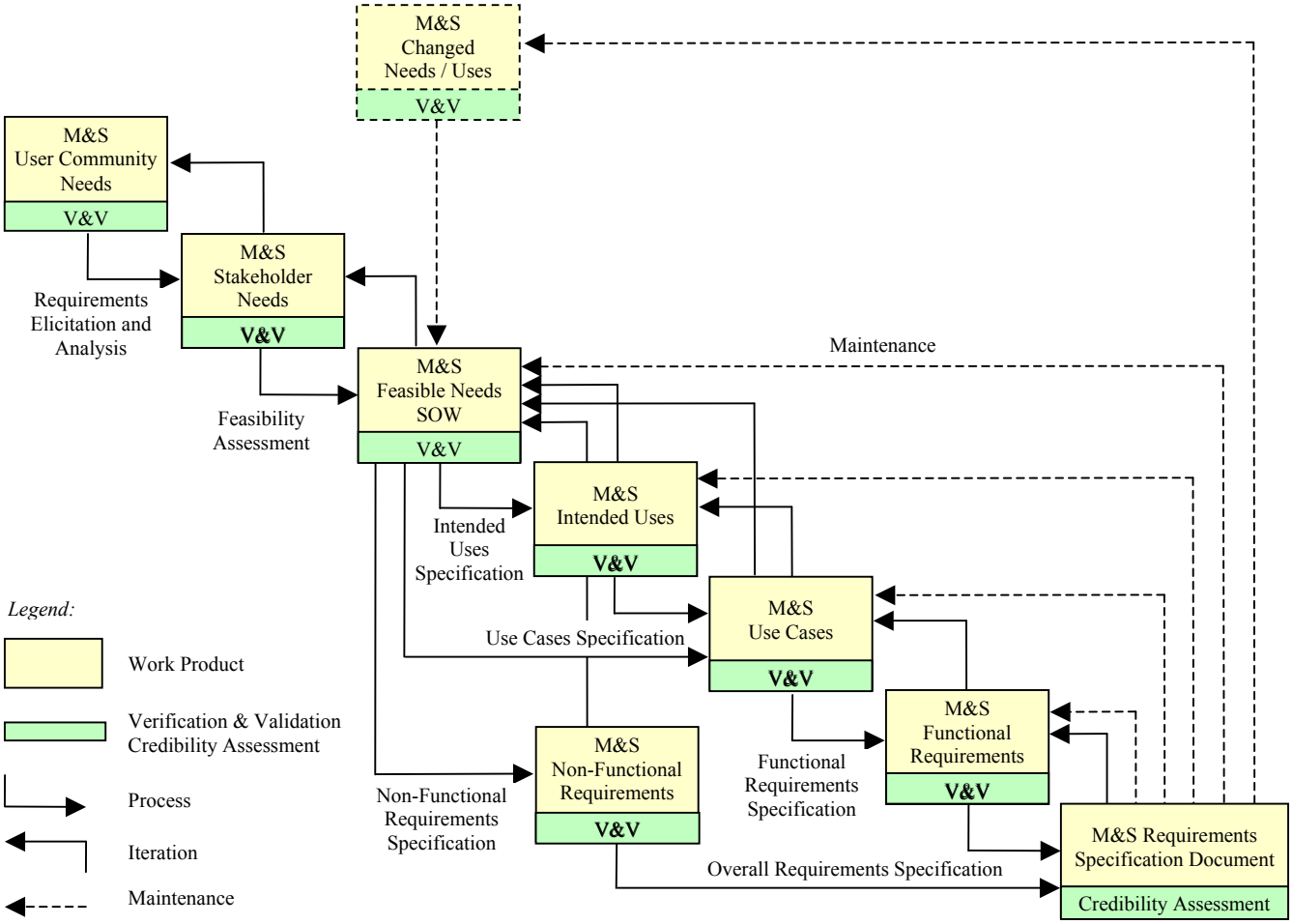

Figure 3: M&S Requirements Engineering Life Cycle

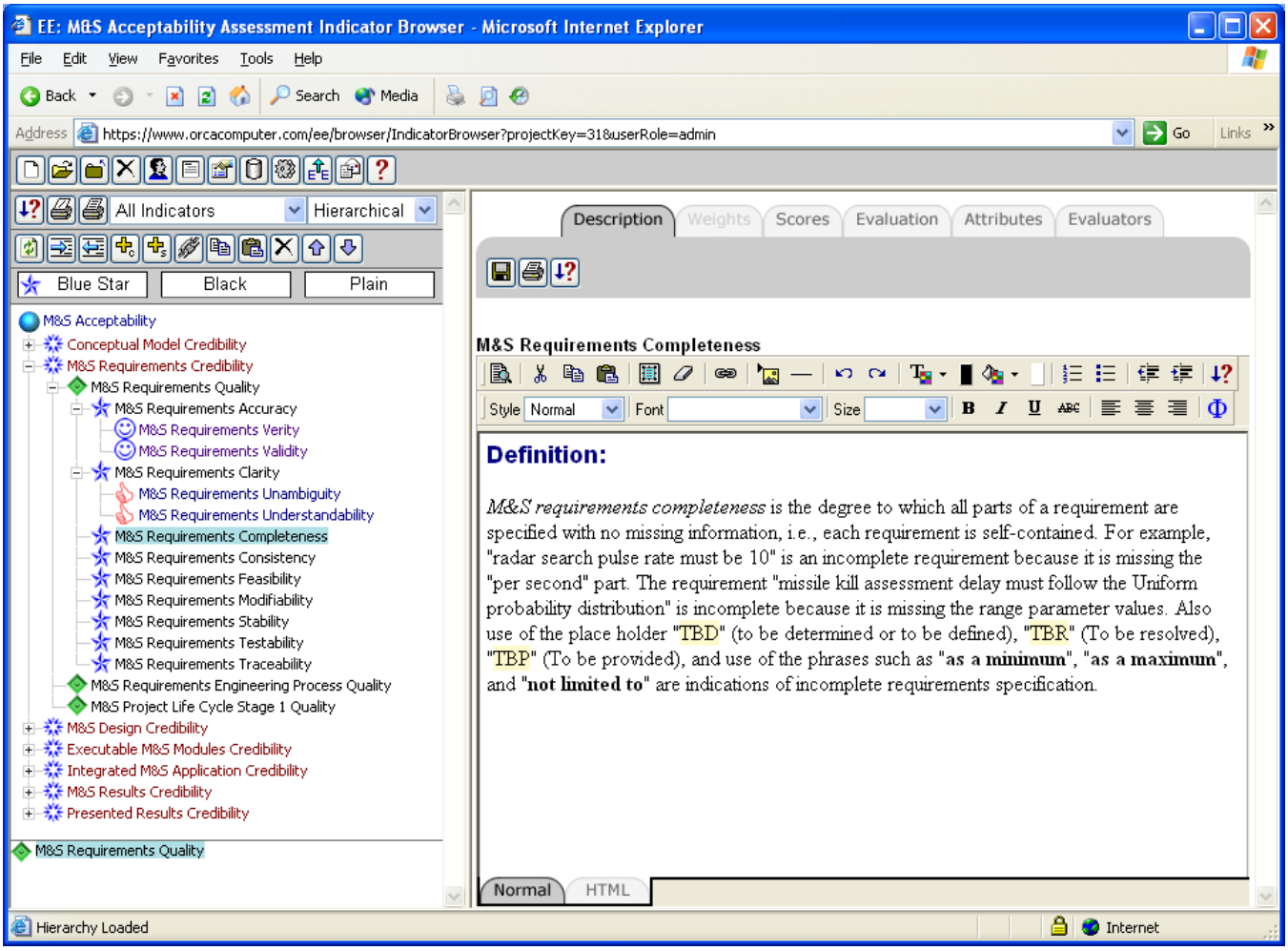

Figure 4: M&S Acceptability Assessment using the Evaluation Environment Web-Based Software System

The following V&V techniques (Balci 1998) can be used for M&S requirements evaluation:

- Desk Checking
- Documentation Checking
- Face Validation
- Inspections
- Reviews
- Walkthroughs

3.3 M&S Design Evaluation Process

This process evaluates the credibility of the M&S design created by conducting the design process based on the M&S requirements specification document. The process integrates the evaluations of: (a) M&S design quality, (b) M&S design process, and (c) M&S project characteristics related to life cycle stage 3.

A reasonably large M&S application design is decomposed into modules to overcome the complexity of development and evaluation. Each M&S module design specification is subjected to evaluation. In addition to assessing design accuracy by performing V&V, quality characteristics such as the following are also assessed: reusability, maintainability, extensibility, interoperability, portability, testability, traceability, understandability, and usability.

EE can be used for collaborative design evaluation as part of the overall M&S acceptability assessment as depicted in Figure 4. A hierarchy of indicators can be created to assess the design quality, design process, and project characteristics. Subject matter experts can be assigned to evaluate leaf indicators. Some leaf indicators can be evaluated by performing V&V, analysis, or direct measurement.

The following V&V techniques (Balci 1998) can be used for M&S design evaluation:

- Audit
- Desk Checking
- Documentation Checking
- Face Validation
- Inspections
- Reviews

- Turing Test
- Walkthroughs
- Cause-Effect Graphing
- Control Analysis
  - Calling Structure Analysis
  - Concurrent Process Analysis
  - Control Flow Analysis
  - State Transition Analysis
- Data Analysis
  - Data Dependency Analysis
  - Data Flow Analysis
- Fault/Failure Analysis
- Interface Analysis
  - Model Interface Analysis
  - User Interface Analysis
- Semantic Analysis
- Structural Analysis
- Symbolic Evaluation
- Syntax Analysis
- Traceability Assessment

3.4 Executable M&S Modules Evaluation Process

This process evaluates the credibility of the executable M&S modules created by conducting the implementation process based on the M&S design specification. The process integrates the evaluations of: (a) executable M&S modules quality, (b) implementation process, and (c) M&S project characteristics related to life cycle stage 4.

An M&S module may be implemented by a team, group, or subcontractor. The implementation process creates executable modules, which can be evaluated by using the following dynamic testing techniques (Balci 1998):

- Acceptance Testing
- Alpha Testing
- Assertion Checking
- Beta Testing
- Bottom-Up Testing
- Comparison Testing
- Compliance Testing
  - Authorization Testing
  - Performance Testing
  - Security Testing
  - Standards Testing
- Debugging
- Execution Testing
  - Execution Monitoring
  - Execution Profiling
  - Execution Tracing
- Fault/Failure Insertion Testing
- Field Testing
- Functional (Black-Box)Testing
- Graphical Comparisons
- Interface Testing
  - Data Interface Testing
  - Model Interface Testing
  - User Interface Testing
- Object-Flow Testing
- Partition Testing
- Predictive Validation
- Product Testing
- Regression Testing
- Sensitivity Analysis
- Special Input Testing
  - Boundary Value Testing
  - Equivalence Partitioning Testing
  - Extreme Input Testing
  - Invalid Input Testing
  - Real-Time Input Testing
  - Self-Driven Input Testing
  - Stress Testing
  - Trace-Driven Input Testing
- Statistical Techniques
- Structural (White-Box)Testing
  - Branch Testing
  - Condition Testing
  - Data Flow Testing
  - Loop Testing
  - Path Testing
  - Statement Testing
- Submodel/Module Testing
- Symbolic Debugging
- Top-Down Testing
- Visualization/Animation

3.5 Integrated M&S Application Evaluation Process

This process evaluates the credibility of the integrated M&S application created by conducting the integration process based on the executable M&S modules. The process integrates the evaluations of: (a) integrated M&S application quality, (b) integration process, and (c) M&S project characteristics related to life cycle stage 5.

One of the principles stated by Balci (1997) dictates that successfully testing each module does not imply overall M&S application credibility. Each module credibility is judged to be sufficient with some error that is acceptable with respect to the project objectives and M&S requirements. We may find each module to be sufficiently credible, but this does not imply that the overall M&S application is sufficiently credible. The allowable errors for the modules may accumulate to be unacceptable for the overall M&S application. Therefore, the integrated overall M&S application must be evaluated even if each module is found to be sufficiently credible.

3.6 M&S Results Evaluation Process

This process evaluates the credibility of the M&S results produced by conducting the experimentation / exercise process based on the integrated M&S application. The process integrates the evaluations of: (a) M&S results quality, (b) M&S experimentation / exercise process, and (c) M&S project characteristics related to life cycle stage 6.

If the M&S application is intended for training purposes, we exercise it typically in a distributed, interactive, and visual manner. If it is used for analysis purposes, we experiment with it to obtain the M&S results in a variety of forms including statistical averages, confidence intervals, graphs, charts, animations, and visualizations.

The design of experiments and statistical analysis of simulation output data are two major areas of discrete event M&S including techniques such as the following: (Banks et al. 2001; Law and Kelton 2000)

- *Response-surface methodologies* can be used to find the optimal combination of parameter values which maximize or minimize the value of a response variable.
- *Factorial designs* can be employed to determine the effect of various input variables on an output variable.
- *Variance reduction techniques* can be implemented to obtain greater statistical accuracy for the same amount of simulation.
- *Ranking and selection techniques* can be utilized for comparing alternative systems.
- *Method of replication*, *method of batch means*, *regenerative method*, and others can be used for statistical analysis of simulation output data.

The experimentation process can be evaluated by using a hierarchy of indicators including the following:

- Are the algorithms used for random variate generation theoretically accurate?
- Are the random variate generation algorithms translated into executable code accurately?
- How well is the random number generator tested?
- Are appropriate statistical techniques implemented to design and analyze the simulation experiments?
- How well are the underlying assumptions satisfied?
- Is the problem of the initial transient (or the start-up problem) appropriately addressed?
- For comparison studies, are identical experimental conditions replicated correctly for each of the alternative operating policies compared?

3.7 Presented Results Evaluation Process

This process evaluates the credibility of the presented results produced by conducting the presentation process based on the produced M&S results. The process integrates the evaluations of: (a) presented results quality, (b) presentation process, and (c) M&S project characteristics related to life cycle stage 7.

The presentation process involves the

- *interpretation* of the M&S results,
- *documentation* of the M&S results, and
- *communication* of the M&S results to the decision makers.

Based on the presented M&S results, the decision makers formulate key decisions including acquisition of a military system, distributing scarce resources, selecting a business strategy, or training military personnel.

A *descriptive model* is a model that describes the behavior of a system without any value judgment on the "goodness" or "badness" of such behavior. All simulation models are descriptive models. Therefore, simulation results must be interpreted. For example, by experimenting with an M&S application, we can estimate the probability of kill as a 95% confidence interval $[0.89 \leq P_{kill} \leq 0.93]$. This M&S result must be interpreted by the analysts to determine if it is a "good" P_{kill} or a "bad" one.

The presentation process also involves the documentation of the M&S results. The documentation quality can be assessed by using a hierarchy of indicators including accessibility, accuracy, completeness, consistency, clarity (unambiguity and understandability), maintainability, portability, and readability.

The communication problem between technical and non-technical people should be recognized and the M&S results should be communicated to the decision makers in an understandable form without any technical jargon.

Due to the complexity of some M&S results, failing to properly interpret, document, and communicate the M&S results may lead to wrong decisions in spite of the fact that the M&S results are sufficiently credible.

4 CONCLUDING REMARKS

The V&V and certification activities must be tied to a well-structured M&S development life cycle. V&V is not a stage but a continuous activity carried out hand in hand with the M&S development throughout the entire life cycle. The use of a well-structured M&S development life cycle is critically important for effectively conducting the V&V and certification activities.

For new M&S application development, concurrent certification is recommended. For certification of an already developed M&S application with or without modifications, some of the evaluation processes can be conducted depending on the M&S Intended Uses. Effective and detailed documentation and the test cases, test data, and test procedures used during development should be provided to facilitate the certification of an already developed M&S application.

Successful certification requires the certification agent to have full access to the M&S application with its associated documentation and data. However, the M&S developer has full control of the M&S application and might not fully cooperate in providing the required material and information to the certification agent. Sometimes, developers view certification as a performance appraisal activity, and they fear that their reputation and potential future funding are at stake if the certification agent identifies problems. Therefore, they sometimes show no desire to cooperate and behave in an adversarial manner against the independent certification agent personnel. The M&S application sponsor has a critical role in resolving this problem (Balci et al. 2002b).

REFERENCES

Balci, O. 1997. Principles of simulation model validation, verification, and testing. *Transactions of the Society for Computer Simulation International* 14 (1): 3-12.

Balci, O. 1998. Verification, validation, and testing. In *The Handbook of Simulation*, ed. J. Banks, 335-393. New York, NY: John Wiley & Sons.

Balci, O. 2001. A methodology for certification of modeling and simulation applications. *ACM Transactions on Modeling and Computer Simulation* 11 (4), 352-377.

Balci, O., R. J. Adams, D. S. Myers, and R. E. Nance. 2002a. A collaborative evaluation environment for credibility assessment of modeling and simulation applications. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 214-220. Piscataway, NJ: IEEE.

Balci, O., R. E. Nance, J. D. Arthur, and W. F. Ormsby. 2002b. Expanding our horizons in verification, validation, and accreditation research and practice. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 653-663. Piscataway, NJ: IEEE.

Balci, O. and W. F. Ormsby. 2000. Well-defined intended uses: an explicit requirement for accreditation of modeling and simulation applications. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 849-854. Piscataway, NJ: IEEE.

Balci, O. and S. D. Saadi. 2002. Proposed standard processes for certification of modeling and simulation applications. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 1621-1627. Piscataway, NJ: IEEE.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2001. *Discrete-Event System Simulation*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

DMSO. 2000. *DoD Verification, Validation, and Accreditation Recommended Practices Guide*, millennium ed. Alexandria, VA: Defense Modeling and Simulation Office.

DoDI. 1996. DoD modeling and simulation verification, validation, and accreditation. Department of Defense Instruction 5000.61, Apr.

IEEE. 1998. *IEEE Standards for Software Verification and Validation.* IEEE Standard 1012. Washington, DC: IEEE Computer Society.

Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. New York, NY: McGraw-Hill.

Orca. 2003. A Web-based Collaborative Evaluation Environment. https://www.orcacomputer.com/ee/

Rae, A., P. Robert, and H.-L. Hausen. 1995. *Software Evaluation for Certification: Principles, Practice, and Legal Liability*. London, UK: McGraw-Hill.

Rational. 2003. Rational Unified Process and Unified Modeling Language. IBM. http://www.rational.com/

Voas, J. M. 1999. Guest editor's introduction: certification: reducing the hidden costs of poor quality. *IEEE Software* 16 (4): 22-25.

AUTHOR BIOGRAPHY

OSMAN BALCI is Professor of Computer Science at Virginia Tech and President of Orca Computer, Inc. He received his Ph.D. degree from Syracuse University in 1981. He currently serves as the Verification, Validation and Accreditation (VV&A) Area Editor of *ACM Transactions on Modeling and Computer Simulation*; Modeling and Simulation (M&S) Category Editor of *ACM Computing Reviews*, and Area Editor of *Simulation: Transactions of the Society for Modeling and Simulation International*. He served as the Editor-in-Chief of two international journals: *Annals of Software Engineering* (1993-2002) and *World Wide Web* (1996-2000). He serves as a member of the Winter Simulation Conference Board of Directors representing the Society for M&S International (SCS) and as a Director at Large for the SCS Board of Directors. Most of Dr. Balci's research has been funded by the Department of Defense since 1983. Since 1998, he has been providing technical services for the Ballistic Missile Defense program. His current areas of expertise center on Software Engineering; E-Systems Engineering (e.g., e-solutions, e-business, e-commerce); M&S; and VV&A, IV&V, and certification of M&S applications and software systems. His email and web addresses are <balci@vt.edu> and <http://manta.cs.vt.edu/balci>.