QUALITY ASSESSMENT, VERIFICATION, AND VALIDATION 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

Many different types of modeling and simulation (M&S) applications are used in dozens of disciplines under diverse objectives including acquisition, analysis, education, entertainment, research, and training. M&S application verification and validation (V&V) are conducted to assess mainly the *accuracy*, which is one of many indicators affecting the M&S application quality. Much higher confidence can be achieved in accuracy if a quality-centered approach is used. This paper presents a quality model for assessing the quality of large-scale complex M&S applications as integrated with V&V. The guidelines provided herein should be useful for assessing the overall quality of an M&S application.

1 INTRODUCTION

Quality is a critically important issue in almost every discipline. Whether we manufacture a product, employ processes or provide services, quality often becomes a major goal. Achieving that goal is the challenge. Many associations have been established worldwide for quality, e.g., American Society for Quality (http://www.asq.org/), Australian Organization for Quality (http://www.aoq.asn.au/), European Organization for Quality (http://www.eoq.org/), and Society for Software Quality (http://www.ssq.org/). Manufacturing companies have quality control departments, business and government organizations have Total Quality Management programs, and software development companies have Software Quality Assurance departments to be able to meet the quality challenge.

The U.S. Department of Defense (DoD) is the largest sponsor and user of Modeling and Simulation (M&S) applications in the world. DoD uses many types of M&S applications (such as continuous, discrete-event, distributed, hardware-in-the-loop, software-in-the-loop, human-in-the-loop, Monte Carlo, parallel, and synthetic environments bringing together simulations and real-world systems) for the purpose of acquisition, analysis or training. DoD M&S applications typically are large-scale and com-

plex and cost millions of dollars to develop over many years, e.g. (MDA 2004). Assuring the quality of diverse types of large-scale and complex M&S applications poses significant technical and managerial challenges.

M&S applications are mostly made up of software or are software based. Software is inherently complex and very difficult to engineer. Under the current technology, we are incapable of developing a reasonably large and complex software product and guaranteeing its 100% accuracy. Accuracy is considered just one of many quality characteristics of an M&S application and is judged by conducting Verification and Validation (V&V). As advocated by Balci et al. (2002), we can increase our confidence in the accuracy of large-scale and complex M&S applications by employing a quality-centered assessment approach.

The purpose of this paper is to present an approach for quality-centered assessment of large-scale and complex M&S applications. Section 2 describes the major indicators of M&S quality. Section 3 presents a quality model for assessing the quality of complex M&S applications. A brief introduction of M&S V&V is provided in Section 4. Concluding remarks are given in Section 5.

2 MAJOR INDICATORS OF M&S QUALITY

M&S application quality is the degree to which the M&S application possesses a desired set of characteristics. Quality assessment is situation dependent since the desired set of characteristics changes from one M&S application to another. M&S application quality is typically assessed by considering the M&S application requirements, intended uses, and project objectives.

Three major indicators influence the M&S application quality as depicted in Figure 1 (Voas 1998). Assessments such as quality assessment, accuracy assessment (i.e., V&V), credibility assessment, and certification assessment are all confidence building activities. We can achieve much better confidence in M&S application quality if we perform the assessment from the perspectives (indicators) of product, process, and project.
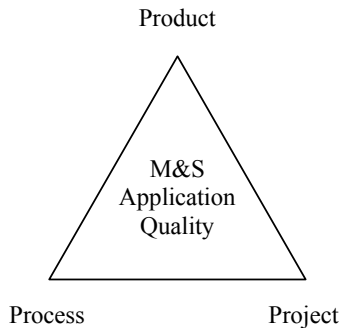
Product

M&S
Application
Quality

Process   Project

Figure 1: M&S Application Quality
Triangle

2.1 M&S Product Quality Assessment

The term *M&S product* implies either (a) the overall completed M&S application, or (b) a work product created during the M&S development life cycle (Balci 2003) such as conceptual model, M&S requirements specification, M&S design specification, and an executable M&S module.

A quality model is presented in Section 3 for assessing the product quality of an overall completed M&S application. The quality model can be applied by geographically dispersed subject matter experts in a collaborative manner by using the Evaluation Environment (EE) software system (Orca 2004) under the evaluation methodology presented by Balci (2001).

The quality of a particular M&S work product can also be assessed by using the EE methodology and software. For each M&S work product, a hierarchy of indicators should be created for assessing the quality of that work product. As an example, we present such a hierarchy of in-

dicators, as depicted in Figure 2, for assessing the product quality of M&S requirements specification. Each quality indicator is defined below.

M&S requirements quality is the degree to which the requirements possess a desired set of characteristics.

M&S requirements accuracy is the degree to which the requirements possess sufficient transformational and representational accuracy.

M&S requirements verity is assessed by conducting M&S requirements verification. M&S requirements verification is substantiating that the M&S requirements are transformed from higher levels of abstraction into their current form with sufficient accuracy judged with respect to the M&S intended uses. M&S requirements verification addresses the question of "Are we creating the M&S requirements <u>right</u>?".

M&S requirements validity is assessed by conducting M&S requirements validation. M&S requirements validation is substantiating that the M&S requirements represent the *real* needs of the application sponsor with sufficient accuracy. M&S requirements validation addresses the question of "Are we creating the <u>right</u> M&S requirements?".

M&S requirements clarity is the degree to which the requirements are unambiguous and understandable. *M&S requirements unambiguity* is the degree to which each statement of the requirements can only be interpreted one way. *M&S requirements understandability* is the degree to which the meaning of each statement of the requirements is easily comprehended by all of its readers.

M&S requirements completeness is the degree to which all parts of a requirement are specified with no missing information, i.e., each requirement is self-contained.
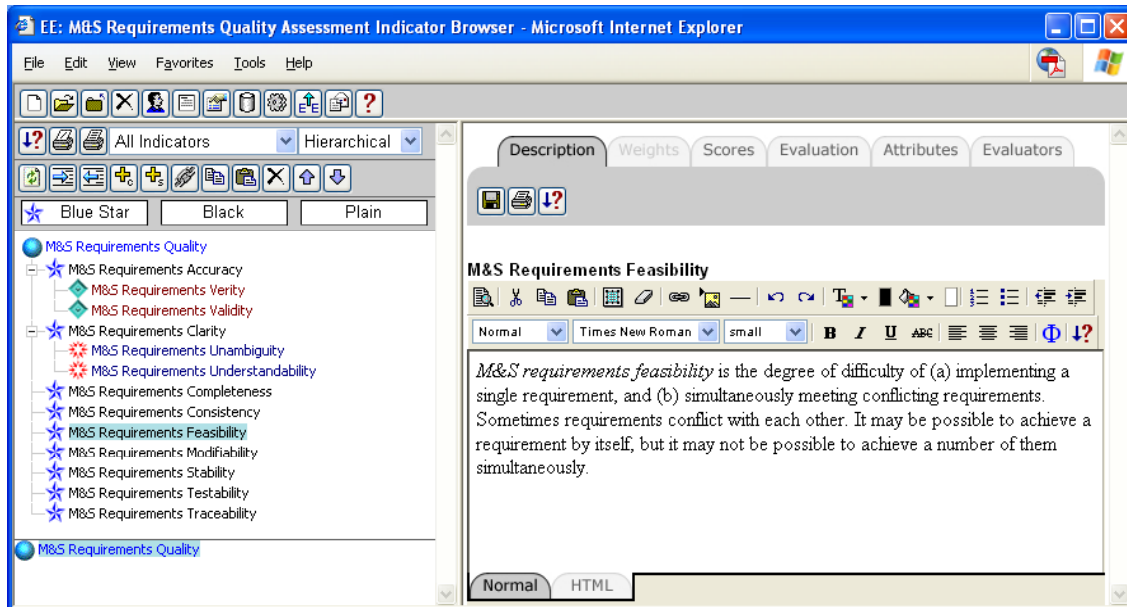
Figure 2: A Hierarchy of Indicators for M&S Requirements Quality Assessment

For example, "radar search pulse rate must be 10" is an incomplete requirement because it is missing the "per second" part. The requirement "missile kill assessment delay must follow the Uniform probability distribution" is incomplete because it is missing the range parameter values. Also use of the place holder "TBD" (to be determined or to be defined), "TBR" (To be resolved), "TBP" (To be provided), and use of the phrases such as "as a minimum", "as a maximum", and "not limited to" are indications of incomplete requirements specification.

M&S requirements consistency is the degree to which (a) the requirements are specified using uniform notation, terminology, and symbology, and (b) any one requirement does not conflict with any other.

M&S requirements feasibility is the degree of difficulty of (a) implementing a single requirement, and (b) simultaneously meeting conflicting requirements. Sometimes requirements conflict with each other. It may be possible to achieve a requirement by itself, but it may not be possible to achieve a number of them simultaneously.

M&S requirements modifiability is the degree to which the requirements can easily be changed.

M&S requirements stability is (a) the degree to which the requirements are changing while the M&S application is under development, and (b) the possible effects of the changing requirements on the project schedule, cost, risk, quality, functionality, design, integration, and testing of the M&S application.

M&S requirements testability is the degree to which the requirements can easily be tested. A testable requirement is the one that is specified in such a way that pass/fail or assessment criteria can be derived from its specification. For example, the following requirement specification is not testable: "The probability of kill should be estimated based on the simulation output data." The following requirement specification is testable: "The probability of kill should be estimated by using a 95% confidence interval based on the simulation output data."

M&S requirements traceability is the degree to which the requirements related to a particular requirement can easily be found. Requirements should be specified in such a way that related requirements are cross-referenced. When it is necessary to change a requirement, those requirements affected by the changed requirement should be easily identified by using the cross references.

2.2 M&S Process Quality Assessment

The term *M&S process* implies a process used to create a work product during the M&S development life cycle (Balci 2003) such as conceptual modeling, requirements, engineering, design, implementation, integration, experimentation, and presentation. The quality of a particular M&S life cycle process can be assessed by creating a hierarchy of indicators and using the EE methodology and software (Balci 2001; Orca 2004).

The manner by which a particular M&S life cycle process is conducted significantly affects the quality of the work product created under that process. Using a structured proven methodology based on industry best practices certainly increases the process quality, which in turn increases the product quality. For example, conducting the M&S requirements engineering process by employing a use case-based approach based on the UML technology versus an ad-hoc approach makes a big difference in the M&S requirements quality. Knowing that the M&S requirements engineering process is properly conducted by using industry best practices increases our confidence in the quality of the M&S process and in the quality of the M&S work product created under that process.

Many indicators influence the M&S process quality including the following (Sommerville 2004):

- Acceptability – the degree to which the defined process is acceptable to and usable by the people who are tasked to execute the process
- Maintainability – the degree to which the process can evolve to reflect changing organizational requirements or identified process improvements
- Quality of the methodology used in executing the process
- Quality of the way the methodology is applied in executing the process
- Rapidity – the degree to which the process can be rapidly executed
- Reliability – the degree to which the process is designed in such a way that process errors are avoided or trapped before they result in product errors
- Robustness – the degree to which the process can be continued in spite of unexpected problems
- Supportability – the degree to which the process activities can be supported by computer-aided software engineering (CASE) tools
- Understandability – the degree to which the process is explicitly defined and easy to comprehend
- Visibility – the degree to which the process activities culminate in clear results so that the process progress is externally visible

2.3 M&S Project Quality Assessment

The third major quality indicator is called project, people, or personnel. This indicator can be decomposed into a hierarchy of indicators including: (a) capabilities or maturity of an organization in developing a large-scale and complex M&S application, (b) quality of application of the or ganization's capabilities, and (c) quality of the people tasked to develop the M&S application.

The Software Engineering Institute at Carnegie Mellon University has led the development of Capability Maturity Model Integration (CMMI) models to provide guidance for creating institutionalized defined processes (SEI 2004). Each CMMI model has two forms of representation: continuous

and staged. The continuous representation depicts the *capabilities* of an organization. The staged representation depicts the *maturity* of an organization. Figure 3 shows the staged representation of the CMMI for Systems Engineering and Software Engineering (CMMI-SE/SW), which is the one that can be used for M&S application development.

Certification at maturity level 3 implies that the organization has institutionalized, defined, and documented all of the process areas listed under levels 2 and 3 in Figure 3 and is ready to execute them for the development of a large-scale and complex M&S application. In addition to being CMMI certified by a credible authority at a contractually required level, the overall M&S application quality is affected by how well the CMMI process areas are executed as well as by the quality of the people employed for the development and management of the M&S application.

3 A QUALITY MODEL FOR M&S APPLICATIONS

A quality model is a characterization of a framework under which the quality is assessed. Based on the Hewlett-Packard's FURPS+ (Grady 1992) software quality model, we propose an M&S application quality model with the hierarchy of indicators as shown in Figure 4. The hierarchy of quality indicators is given for the higher levels of the quality model. The leaf indicators need to be further decomposed into other indicators until the leaf indicators are directly assessable. The hierarchy forms an acyclic graph where an indicator may influence more than one parent indicator such as the Accuracy indicator influencing Reliability and Functionality in Figure 4.

Some of the quality indicators given in Figure 4 may not be applicable for some M&S applications. Many different types of M&S application exist such as continuous, discrete-event, distributed, hardware-in-the-loop, software-in-the-loop, human-in-the-loop, Monte Carlo, parallel, and synthetic environments bringing together simulations and real-world systems. An M&S application may be used for, many different purposes including acquisition, analysis education, entertainment, research, and training. The M&S application type and the purpose for which it is constructed should be taken into consideration in modifying and expanding the hierarchy of indicators given in Figure 4.

Maturity Level 5: Optimizing
Organizational Innovation and Deployment
Causal Analysis and Resolution

Maturity Level 4: Quantitatively Managed
Organizational Process Performance
Quantitative Project Management

Maturity Level 3: Defined
Requirements Development
Technical Solution
Product Integration
Verification
Validation
Organizational Process Focus
Organizational Process Definition
Organizational Training
Integrated Project Management
Risk Management
Decision Analysis and Resolution

Maturity Level 2: Managed
Requirements management
Project planning
Project Monitoring and Control
Supplier Agreement Management
Measurement and Analysis
Process and Product Quality Assurance
Configuration Management
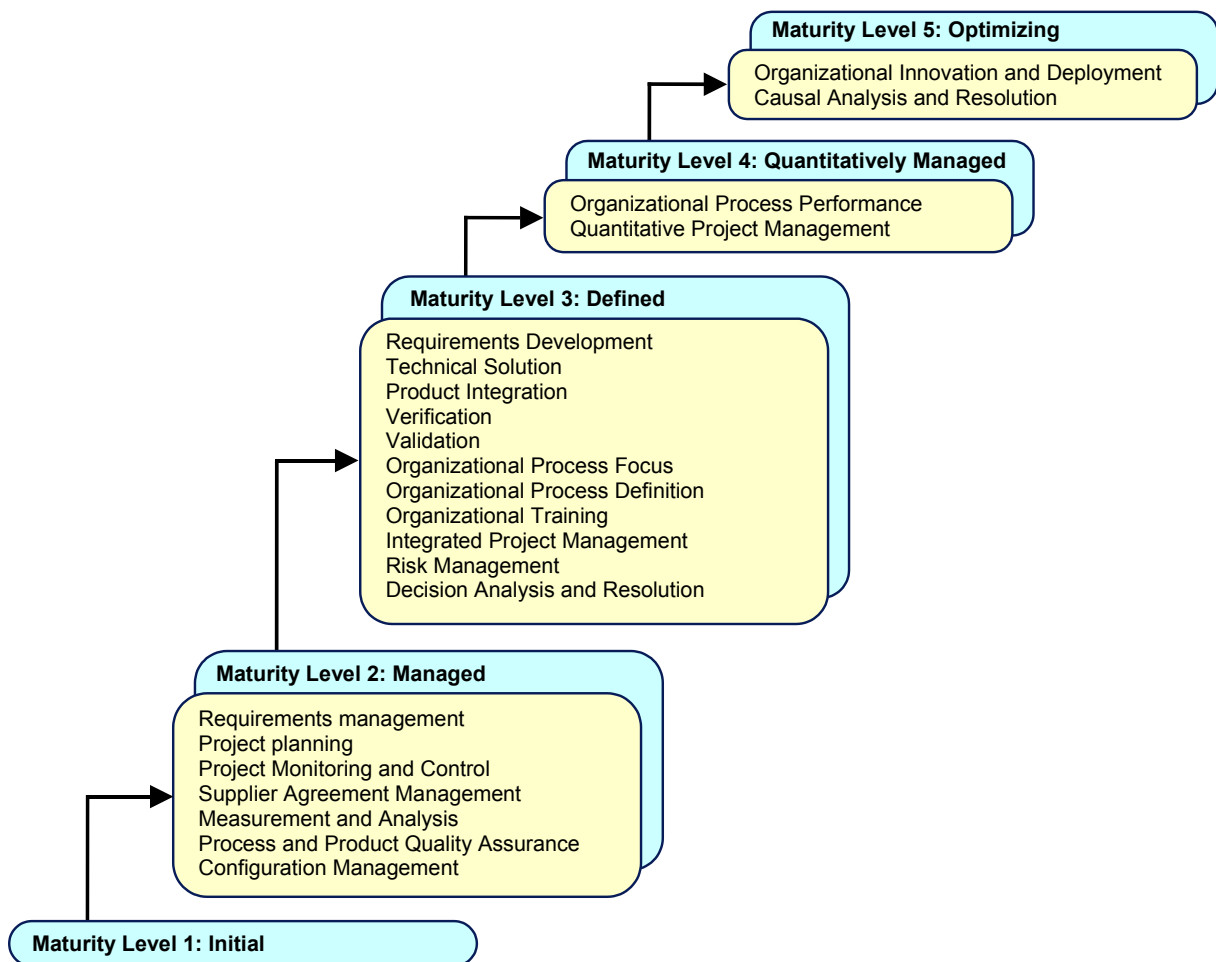
Maturity Level 1: Initial
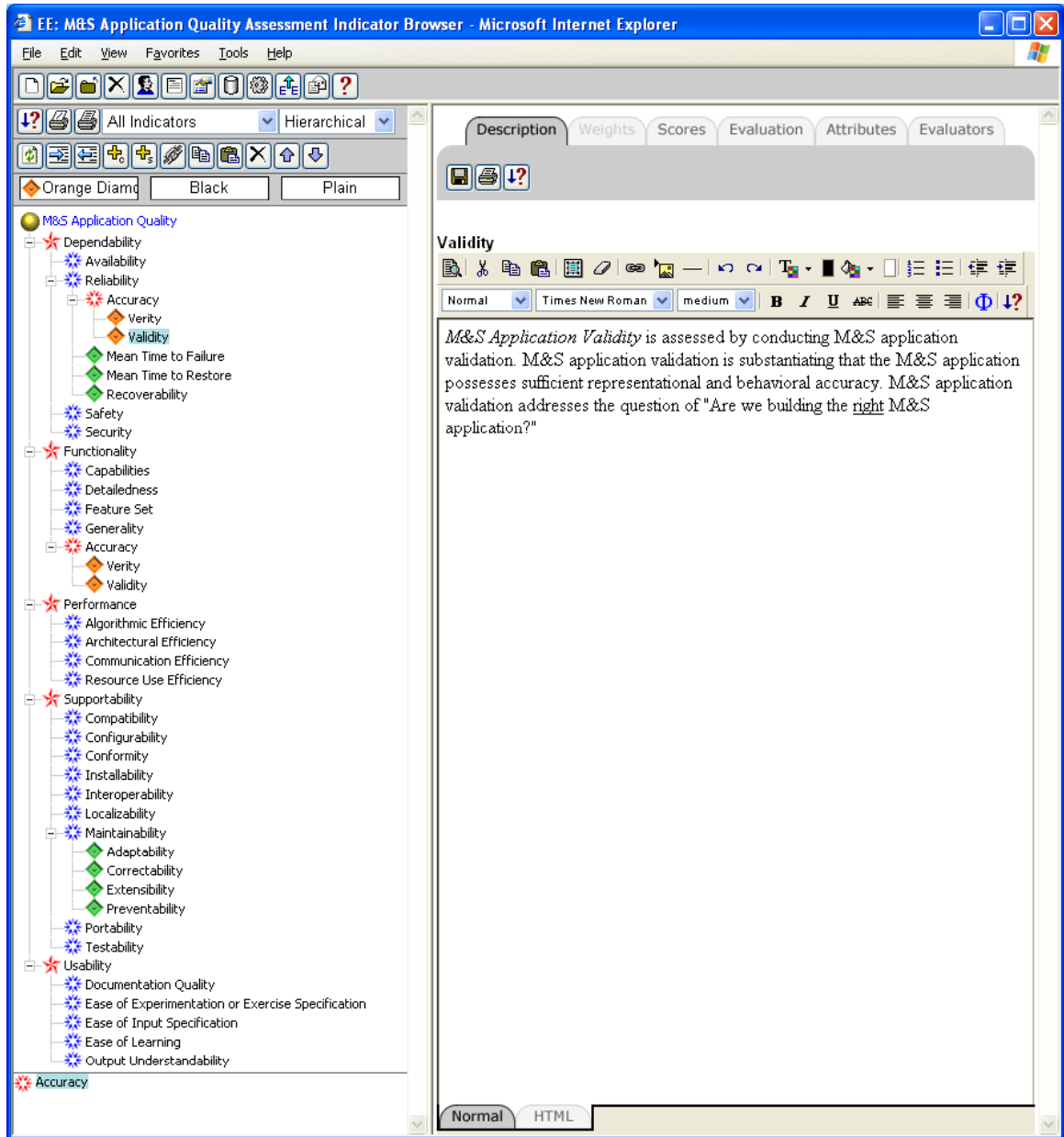
Figure 3: CMMI-SE/SW Staged Process Areas

Figure 4: M&S Application Quality Model Hierarchy of Indicators in an EE Project

Definitions of the indicators of the quality model are given below (Pressman 2005; Sommerville 2004).

M&S Application Dependability is the degree to which the M&S application (a) delivers services when requested, (b) delivers services as specified, (c) operates without catastrophic failure, and (d) protects itself against accidental or deliberate intrusion.

M&S Application Availability is the probability that the M&S application functions according to its requirements at a given point in time. Availability refers to the ability of the M&S application to deliver services when requested.

M&S Application Reliability is the degree to which the M&S application performs its required functions without

failure under prescribed conditions in a specified period of time for a specific purpose. M&S application reliability refers to the ability of the M&S application to deliver services as specified.

M&S Application Accuracy is the degree to which the M&S application possesses sufficient transformational and representational/behavioral accuracy.

M&S Application Verity is assessed by conducting M&S application verification, which is substantiating that the M&S application is transformed from one form into another with sufficient accuracy. M&S application verification addresses the question of "Are we building the M&S application <u>right</u>?"

M&S Application Validity is assessed by conducting M&S application validation, which is substantiating that the M&S application possesses sufficient representational and behavioral accuracy. M&S application validation addresses the question of "Are we building the <u>right</u> M&S application?"

M&S Application Mean Time to Failure (MTTF) is the average time between observed M&S application failures. MTTF = 300 hours means that, on the average, one failure can be expected to occur every 300 hours.

M&S Application Mean Time to Restore (MTTR) is the average time it takes to restore the M&S application after failure.

M&S Application Recoverability is the degree to which the M&S application provides mechanisms to enable users to recover from errors.

M&S Application Safety is the ability of the M&S application to operate, normally or abnormally, without threatening people or the environment. M&S safety refers to the ability of the M&S application to operate without catastrophic failure. The safety may be an issue particularly for training simulations.

M&S Application Security is the degree to which the M&S application provides protection and authentication of information in transit or stationary, as well as the confidentiality of sensitive information. M&S security refers to the ability of the M&S application to protect itself against accidental or deliberate intrusion.

M&S Application Functionality is the degree to which the M&S application completely captures all of the desired functional modules that need to be present.

M&S Application Capabilities is the degree to which the M&S application is capable of performing its feature set, e.g., capability of simulating a particular combat at the soldier level of granularity.

M&S Application Detailedness is the degree to which the M&S application is characterized by abundant use of detail or thoroughness of treatment.

M&S Application Feature Set is the degree to which the M&S application provides the set of features that need to be present, e.g., simulating a particular combat.

M&S Application Generality is the degree to which the M&S application can be used for a wide range of intended uses.

M&S Application Performance is the degree to which the M&S application executes its work in a speedy, efficient, and productive manner.

M&S Application Algorithmic Efficiency is the degree to which the algorithms used in the M&S application provide the optimal execution time.

M&S Application Architectural Efficiency is the degree to which the M&S application architecture enables the optimal execution time.

M&S Application Communication Efficiency is the degree to which the M&S application fulfills its purpose of communicating with its user over a network without waste of resources. Communication efficiency is influenced by the communication protocol (e.g., HTTP or RMI) used by the M&S application, encryption/decryption of the communication, or the existence of a firewall.

M&S Application Resource Use Efficiency is the degree to which the M&S application fulfills its purpose without waste of resources such as CPU, main memory, and hard disk space.

M&S Application Supportability is the degree to which the M&S application can be supported.

M&S Application Compatibility is the degree to which the M&S application can be integrated into or used with other M&S applications, products, or systems.

M&S Application Configurability is the degree to which the M&S application can easily be set up or configured for a particular application or intended use.

M&S Application Conformity is the degree to which the M&S application adheres to standards and conventions.

M&S Application Installability is the degree to which the M&S application can easily be prepared for use.

M&S Application Interoperability is the degree to which the M&S application in a distributed environment (e.g., federation of models) can exchange data with one or more other M&S applications and be able to use the data that has been exchanged.

M&S Application Localizability is the degree to which the M&S application can easily be adopted, preferably via preferences or options, (a) to satisfy the needs of languages other than English, and (b) to local standards such as decimal separator, currency symbol, time zone, calendar, etc.

M&S Application Maintainability is the degree to which the M&S application facilitates changes for:

- adaptations required as the M&S application's external environment evolves (*adaptive maintenance*),
- fixing bugs and making corrections (*corrective maintenance*),
- enhancements brought about by changing customer requirements (*perfective maintenance*), and
- preventing potential problems or for reengineering (*preventive maintenance* or *software reengineering*).

M&S Application Adaptability is the degree to which the M&S application can accommodate changes to its external environment.

M&S Application Correctability is the degree to which the M&S application facilitates changes for fixing bugs and making corrections.

M&S Application Extensibility is the degree to which the M&S application capabilities can be extended by modifying current features or adding new features.

M&S Application Preventability is the degree to which the M&S application facilitates changes for preventing potential problems or for reengineering.

M&S Application Portability is the degree to which the M&S application can easily be transformed to run on more than one hardware or software environment.

M&S Application Testability is the degree to which the M&S application facilitates the creation of test criteria and conducting tests to determine whether those criteria have been met.

M&S Application Usability is the degree to which the M&S application can easily be employed for its intended use.

M&S Application Documentation Quality is the degree to which the M&S application external documentation (e.g., user manuals, reference guides, online help) possesses a desired set of characteristics.

Ease of Experimentation or Exercise Specification is the degree to which a simulation experiment (for analysis) or a simulation exercise (for training) can easily be specified.

Ease of M&S Application Input Specification is the degree to which the input conditions and data of the M&S application are easily specified under a set of prescribed intended uses.

M&S Application Ease of Learning is the ease with which the M&S application can be learned.

M&S Application Output Understandability is the degree to which the meaning of the M&S application output is easily comprehended by its users under a set of prescribed intended uses.

4   M&S VERIFICATION AND VALIDATION

M&S V&V typically focuses on the assessment of M&S transformational accuracy (verification) and representational/behavioral accuracy (validation). The quality model in Figure 4 shows the M&S accuracy being evaluated by the M&S verity and M&S validity indicators. M&S verity is assessed by conducting M&S verification and M&S validity is assessed by conducting M&S validation.

Undoubtedly, accuracy is the most important quality characteristic of an M&S application. V&V is used to assess the accuracy quality characteristic. However, under the current state of the art, we are unable to claim a level of accuracy of a reasonably large and complex M&S application *with 100% confidence* due to many reasons including M&S complexity, reliance on human judgment, qualitative measurements, lack of data, and lack of exhaustive testing. Hence, M&S V&V is viewed as a "confidence building" activity (Balci et al. 2002).

For a reasonably large and complex M&S application, the "confidence building" activity must be performed by assessing not only the M&S accuracy, but also other M&S quality characteristics such as the ones given in the quality model in Figure 4. Successful assessment of the overall M&S application quality increases our confidence in M&S accuracy.

The V&V activities must be tied to a well-structured M&S development life cycle (Balci 2003). 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 "confidence building" activities.

More than 100 V&V techniques exist. Many of them are described by Balci (1998). Balci (2003) identifies the V&V techniques applicable for use during different stages of the M&S application development life cycle. Each V&V technique provides information from a different perspective so as to increase our confidence in the M&S accuracy. The more V&V techniques are applied the more information is provided and the more confidence is gained in the M&S accuracy.

5   CONCLUDING REMARKS

A set of *acceptability criteria* should be specified for each *intended use* (Balci and Ormsby 2000) of an M&S application in the legal contract signed by the M&S sponsor and developer. *Acceptance testing* (Balci 1998) is typically conducted by an independent third party, e.g., M&S Certification Agent, to judge if the M&S application meets the prescribed acceptability criteria. The acceptance testing commonly includes the assessment of many quality characteristics as well as the accuracy.

Judging if the M&S application satisfies the acceptability criteria is a "confidence building" activity. The conclusion of the acceptance testing should also provide a confidence level such as "*We are 95% confident* that the M&S application satisfies the acceptability criteria for the stated intended use." However, it is usually very difficult to numerically define the confidence level; therefore, nominal characterizations such as "exceptionally", "exceedingly", "highly", "remarkably", or "notably" is used in place of the "95%" numerical value.

The confidence level at which the acceptance testing is to be concluded should be dictated by the M&S application sponsor. The confidence level can be more convincingly substantiated if a quality-centered approach is used in performing the acceptance testing. CMMI certification of the M&S application developer and the quality of the people tasked to develop the M&S application significantly affect the confidence level justification.

REFERENCES

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. 2003. Verification, validation, and certification of modeling and simulation applications. In *Proceedings*

of the 2003 Winter Simulation Conference*, ed. S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 150-158. Piscataway, NJ:IEEE.

Balci, O., R. E. Nance, J. D. Arthur, and W. F. Ormsby. 2002. 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.

Grady, R. B. 1992. Practical Software Metrics for Project Management and Process Improvement. NJ : Prentice-Hall.

MDA. 2004. Missile Defense Agency Web site. `http://www.acq.osd.mil/bmdo/`

Orca. 2004. A Web-based Collaborative Evaluation Environment. `https://www.orcacomputer.com/ee/`

Pressman, R.S. 2005. Software Engineering: A Practitioner's Approach, 6th ed. New York, NY:McGraw-Hill.

SEI. 2004. CMMI (Capability Maturity Model Integration) Web site. Software Engineering Institute, Carnegie Mellon University. `http://www.sei.cmu.edu/cmmi/`

Sommerville, I. 2004. *Software Engineering*. 7th ed. Reading, MA: Addison-Wesley.

Voas, J. M. 1998. The Software Quality Certification Triangle. *CrossTalk – The Journal of Defense Software Engineering* 11 (11): 12-14.

AUTHOR BIOGRAPHY

OSMAN BALCI is Professor of Computer Science at Virginia Tech. 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*; Area Editor of *Simulation: Transactions of the Society for Modeling and Simulation International*, and an Editorial Board Member of the *International Journal of Simulation and Process Modeling*. 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; M&S; and VV&A, IV&V, and certification of M&S applications and software systems. His e-mail address is `<balci@vt.edu>` and his web address is `<http://manta.cs.vt.edu/balci>`.