Golden Rules of Verification, Validation, Testing, and Certification of Modeling and Simulation Applications

Osman Balci

Department of Computer Science
Virginia Polytechnic Institute and State University (Virginia Tech)
Blacksburg, Virginia 24061, U.S.A.

Abstract

The ever-increasing complexity of the problems we try to solve using modeling and simulation (M&S) poses significant technical challenges for substantiating the sufficient accuracy and certification of M&S applications. This paper presents 20 golden rules to guide an M&S practitioner in conducting verification, validation, testing, and certification. Proper application of these golden rules increases the probability of success in establishing the sufficient credibility of an M&S application.

1   Introduction

A *model* is a representation and abstraction of something such as an entity, a real system, a proposed system design or an idea. *Simulation* is the act of experimenting with or exercising a model or a number of models under diverse objectives such as problem solving, training, acquisition, entertainment, research and education. Since modeling is an integral part of a simulation, we call the entire development process as *modeling and simulation* (M&S) and the end product as *M&S Application*.

Many types of M&S applications exist such as discrete, continuous, Monte Carlo, system dynamics, gaming-based, agent-based, virtual reality, distributed, web-based, hardware-in-the-loop, software-in-the-loop, and human-in-the-loop [ACM 2010]. Each M&S application type poses its own technical challenges for credibility assessment.

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* <u>right</u>?"

- *x Validation* deals with the assessment of *behavioral or representational accuracy* of the *x* and addresses the question of "Are we creating the <u>right</u> *x*?"

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

Testing is the process of designing a test, specifying test conditions and test data, and determining a procedure to follow for the purpose of judging transformational accuracy (verity) and/or representational/behavioral accuracy (validity). Testing is conducted to perform verification and/or validation. We refer to the entire accuracy assessment activity as Verification, Validation, and Testing (VV&T).

The International Organization for Standardization (ISO) defines certification and accreditation as follows: [Rae, Robert, and Hausen 1995]

Certification is a "procedure by which a third party gives written assurance that a product, process or service conforms to specified characteristics."

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 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 sponsor as providing some form of guarantee of quality and credibility. Based on that 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.

2   The Golden Rules

A *golden rule* is a fundamental principle that increases the probability of success, if followed properly. The golden rules are described below in no particular order.

2.1   Golden Rule 1

Model and/or Simulation (M/S) VV&T should be conducted hand in hand as integrated within the entire M&S application development life cycle.

VV&T is not a stage or step in the M&S development life cycle, but a continuous activity throughout the entire life cycle. Accuracy is not something that can be imposed upon after the fact; it has to be assessed while the work is being performed.

2.2   Golden Rule 2

M/S VV&T outcome should not be considered as a binary variable where M/S accuracy is either perfect or totally imperfect.

Since a model, by definition, is an abstraction, perfect representation is never expected. Therefore, M/S accuracy is not assessed to conclude with a binary decision, where 1 implies "perfectly accurate" and 0 implies "totally inaccurate." M/S accuracy should be judged on a scale defined by nominal scores such as excellent, very good, satisfactory, marginal, deficient, and unsatisfactory.

2.3   Golden Rule 3

An M/S is built for a prescribed set of intended uses and its accuracy is judged with respect to those intended uses.

A model, by definition, is a representation and as such the representation can be created in many different ways depending on the objectives for which the model is intended for use. The intended uses of an M&S application dictate how representative the M/S should be. Sometimes, the nominal score "Very Good" M/S accuracy may be sufficient; sometimes, "Excellent" accuracy may be required depending on the criticality of the decisions to be made based on the M&S application results. Therefore, M/S accuracy should be judged with respect to a predefined set of intended uses, for which the M/S is created.

The adjective "sufficient" should be used in front of terms such as accuracy, verity, validity, quality, and credibility, to indicate that the judgment is made with respect to the prescribed set of intended uses. It is more appropriate to say "the model is sufficiently valid" than saying "the model is valid." Here "sufficiently valid" implies that the validity is judged with respect to the prescribed set of intended uses and found to be sufficient.

2.4   Golden Rule 4

M/S VV&T requires independence to prevent developer's bias.

M/S VV&T is meaningful when conducted in an independent manner by an unbiased person, group, or agent who is independent to the M&S application developer. The people involved in M&S application development may be biased when it comes to VV&T, because they may fear that negative VV&T results may be used for their performance appraisal.

M/S VV&T should be conducted under true independence requiring technical, managerial, and financial independence [IEEE 1998]. *Technical Independence* implies that the M/S VV&T agent determines, prioritizes, and schedules its own tasks and efforts. *Managerial Independence* implies that the M/S VV&T agent

reports to the M&S application sponsor independently of the developer organization. *Financial Independence* implies that the M/S VV&T agent is allocated its own budget and does not rely on M&S application development budget.

2.5   Golden Rule 5

M/S VV&T is difficult and requires creativity and insight.

M/S VV&T is difficult due to many reasons including lack of data, lack of sufficient problem domain-specific knowledge, lack of qualified subject matter experts, many qualitative elements to assess, and inability to effectively employ M/S developers due to their conflicts of interest. Designing an effective test, identifying test cases, and developing a test procedure require creativity and insight. VV&T experience is required to be able to determine which of the more than 100 V&V techniques [Balci 1998] are most effective for a given V&V task.

2.6   Golden Rule 6

M/S VV&T is situation dependent.

VV&T is applied depending on the particular accuracy assessment task, M/S type, M/S size, M/S complexity, and the nature of the artifact subjected to VV&T. A number of most effective VV&T techniques for one VV&T situation may not be so for another. The VV&T approach, techniques, and tools should be selected depending on the VV&T task at hand.

2.7   Golden Rule 7

M/S accuracy can be claimed only for the intended uses for which the M/S is tested.

M/S accuracy is assessed using VV&T for a particular intended use, under which the M/S input conditions are defined. The M&S application that works for one set of input conditions under a given intended use may produce absurd output when conducted under another set of input conditions. For example, assume that an M&S application is developed

for the intended use of finding the best light timing for a traffic intersection during the evening rush hour (input conditions). Assume that the M&S application is found to be sufficiently accurate for such use under the stated input conditions. Sufficient accuracy of that M&S application cannot be assumed for other input conditions such as morning rush hour, noon rush hour, or during the entire day of operation.

2.8   Golden Rule 8

Complete testing is not possible for large and complex models and/or simulations.

"The only exhaustive testing there is, is so much testing that the tester is exhausted!" Exhaustive (complete) testing requires that the M/S is tested under all possible input values. Combinations of feasible values of model input variables can generate millions of logical paths in the execution of a large and complex simulation model. Due to time and budgetary constraints, it is impossible to test the accuracy of millions of logical paths.

When using test data, it should be noted that the "law of large numbers" simply does not apply. The question is not how much test data are used, but what percentage of the potential model input domain is covered by the test data. The higher the percentage of coverage the higher the confidence we can gain in model accuracy.

2.9   Golden Rule 9

M/S VV&T activities should be considered as confidence building activities.

We are unable to claim sufficient accuracy of a reasonably large and complex M/S with 100% confidence due to M/S complexity, lack of data, reliance on qualitative human judgment, and lack of complete testing. The M/S VV&T activities are conducted until sufficient confidence in M/S accuracy is gained. Therefore, M/S VV&T activities should be viewed as "confidence building" activities.

Accuracy is certainly the most important quality indicator and VV&T is conducted to assess it. However, for a large and complex

M&S application, we are unable to substantiate sufficient accuracy with 100% confidence. In this case, assessment of other quality indicators helps us increase our confidence in sufficient accuracy of the M&S application.

2.10 Golden Rule 10

M/S VV&T activities should be planned and documented throughout the entire M&S development life cycle.

The M/S VV&T activities should not be conducted in an *ad hoc* fashion. Planning such activities is required for (a) scheduling VV&T tasks throughout the entire M&S application development life cycle, (b) identifying software tools to acquire, (c) identifying methodologies and techniques to use, (d) assigning roles and responsibilities, and (e) allocating resources such as personnel, facilities, tools, and finances.

All VV&T activities should be documented for certification, regression testing, re-testing, and re-certification. All of the VV&T artifacts such as test designs, test data, test cases, and test procedures should be documented and preserved for re-use during the maintenance stage of the M&S application life cycle.

2.11 Golden Rule 11

Errors should be detected as early as possible in the M&S application development life cycle.

M&S application development must start with problem formulation and must be carried out process by process in an orderly fashion in accordance with a comprehensive blueprint of development (i.e., life cycle). Skipping the early stages of development and jumping into programming results in an approach called "build-and-fix" and must be avoided.

Detection and correction of errors as early as possible in the M&S application development life cycle results in reduced development time and assures better quality. Some vital errors may not be detectable in later stages of the life cycle due to increased, unmanageable complexity. It is relatively easier to detect, localize and correct errors in an incremental manner as the M/S development progresses.

2.12 Golden Rule 12

Double validation problem should be recognized and resolved properly.

A typical validation test is conducted by running the simulation model with the "same" input data that drive the system, and then comparing the model and system outputs to determine how similar they are. The amount of correspondence between the model and system outputs is examined to judge the validity of the model.

However, in conducting this validation test, another validation test should be recognized and performed before this test. That validation test deals with substantiating that the model and system inputs match each other with sufficient accuracy. This test is also referred to as *input data model validation*, which must be successfully performed before the model validation test.

2.13 Golden Rule 13

Successfully testing each submodel (module) does not imply overall model validity.

Submodels representing subsystems can be tested individually. Since a model or submodel, by definition, is an abstraction, perfect representation is never expected and some representation error is allowed. Each submodel can be found to be acceptable with respect to the intended uses with some tolerable error in its representation. However, the allowable errors for the submodels may accumulate to be unacceptable for the whole model. Therefore, the whole model must be tested even if each submodel is individually found to be acceptable.

2.14 Golden Rule 14

Formulated problem accuracy greatly affects the acceptability and credibility of M&S results.

It has been said that "a problem correctly formulated is half solved" or "proper formulation of a problem is 50% of its solution." The M&S life cycle starts with problem formulation. Based on the formulated problem, the system or domain containing the problem is

defined and its characteristics are identified. Based on the defined problem domain, M&S application requirements are engineered and the requirements become the point of reference for the M&S application development throughout the rest of the life cycle. An incorrectly defined problem results in simulation results that are irrelevant. Formulated problem accuracy greatly affects the credibility and acceptability of simulation results. Sufficient time and effort must be expended to properly define the problem.

2.15 Golden Rule 15

Type I, II and III errors should be recognized and prevented.

Three types of errors may be committed in developing an M&S application:

- *Type I Error* is committed when the M&S application results are rejected when in fact they are sufficiently credible.

- *Type II Error* is committed when the M&S application results are accepted when in fact they are not sufficiently credible.

- *Type III Error* occurs when the wrong problem is solved and is committed when the formulated problem does not completely contain the actual problem.

Committing Type I Error unnecessarily increases the M&S application development cost. The consequences of Type II and Type III Errors can be catastrophic especially when critical decisions are made on the basis of M&S application results. Type III Error implies that the problem solution and the simulation results will be irrelevant when it is committed.

Two types of risks are recognized [Balci and Sargent 1981]:

- *Model Builder's Risk:* probability of committing Type I Error

- *Model User's Risk:* probability of committing Type II Error

The M/S VV&T activities should focus on minimizing these risks as much as possible.

2.16 Golden Rule 16

Certification should be conducted by an independent third party.

Certification is meaningful if and only if it is conducted in a truly independent manner. True independence requires technical, managerial, and financial independence [IEEE 1998].

- *Technical Independence* implies that the certification agent (third party) 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 M&S 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.

2.17 Golden Rule 17

Certification should be conducted concurrently throughout the entire development life cycle of a new M&S application.

For new M&S application development, certification should be conducted concurrently hand in hand with the development activities. For certification of an already developed M&S application with or without modifications, effective and detailed documentation, test cases, test data, and test procedures used during development should be provided to facilitate the certification. [Balci 2001]

2.18 Golden Rule 18

A certification agent should be accredited.

A company or organization interested in serving as M&S certification agent applies to an accreditation authority, which examines the acceptability and maturity of the applicant's certification processes and the qualifications of

the key people who 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 certification processes and provide certification which is unbiased, fair, cost effective, and consistent.

2.19 Golden Rule 19

M&S application sponsor should clearly dictate the rules of conduct between the M&S application developer and M&S application certification agent.

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.* 2002].

2.20 Golden Rule 20

Certification outcome should be presented with a level of confidence.

Certification outcome should not be a binary decision, where 1 implies "certified" and 0 implies "not certified." Certification outcome should be reached with a *confidence level* expressed as a nominal value such as very low, low, average, high, and very high.

Certification may not be carried out at a desired level of quality due to many factors including lack of data, schedule delays, loss of resources, changing requirements, and development refocus. The level of quality at which certification is conducted influences the

level of confidence with which a certification outcome is reached as depicted in Figure 1.
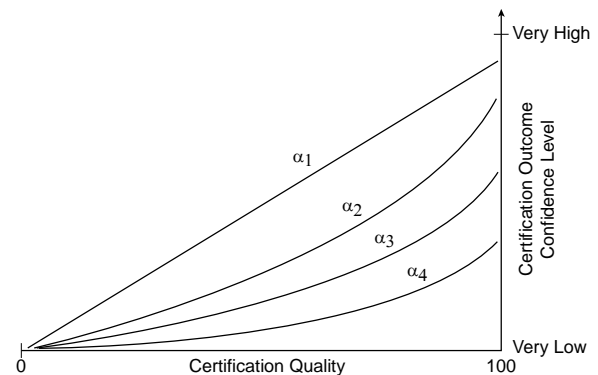
Figure 1. Certification quality versus certification outcome confidence level

As the certification quality increases so does our confidence in reaching the certification outcome. The relationship between certification quality and certification outcome confidence level is situation dependent and is notionally illustrated by curves with different α shape parameter values as shown in Figure 1.

3   Concluding Remarks

The M/S VV&T and certification activities can be conducted under the guidance of the methodology presented by Balci [2001]. This methodology enables the decomposition of a complex assessment problem into a hierarchy of indicators in the form of an acyclic graph.

How much to test or when to stop testing depends upon how much confidence is needed with respect to the project objectives and M&S application requirements. The testing should continue until we achieve sufficient confidence in credibility and acceptability of M&S application results. The sufficiency of the confidence is dictated by the intended uses.

An M&S development organization should establish an M&S quality assurance (QA) group or department so as to gain some independence for the conduct of VV&T activities. The QA department should report to the upper management independently of the development department so that some internal independence is achieved. QA should go beyond VV&T and should also be responsible for assessing other

M/S quality characteristics such as performance, interoperability, maintainability, reusability, portability, and usability.

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

References

ACM (2010), "ACM SIGSIM M&S knowledge repository," Special Interest Group on Simulation and Modeling (SIGSIM), Association for Computing Machinery (ACM), New York, NY, http://www.acm-sigsim-mskr.org

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

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. 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*, IEEE, Piscataway, NJ, pp. 653-663.

Balci, O. and R. G. Sargent (1981), "A methodology for cost-risk analysis in the statistical validation of simulation models," *Communications of the ACM 24*, 4 (Apr.), 190-197.

IEEE (1998), "IEEE standards for software verification and validation," IEEE Standard 1012. IEEE Computer Society, Washington, DC.

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

Author Biography

Osman Balci is a Professor of Computer Science at Virginia Polytechnic Institute and State University (Virginia Tech). He received B.S. and M.S. degrees from Boğaziçi University (Istanbul) in 1975 and 1977, and M.S. and Ph.D. degrees from Syracuse University (New York) in 1978 and 1981. Dr. Balci currently serves as an Area Editor of *ACM Transactions on Modeling and Computer Simulation*, Modeling and Simulation (M&S) Category Editor of *ACM Computing Reviews*, and Editor-in-Chief of *ACM SIGSIM M&S Knowledge Repository*. He served as the Editor-in-Chief of two international journals: *Annals of Software Engineering*, 1993-2002 and *World Wide Web*, 1996-2000. He served as Chair of ACM SIGSIM, 2008-2010; and Director at Large of the Society for M&S International, 2002-2006. He can be contacted at balci@vt.edu and http://manta.cs.vt.edu/balci.