Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

EXPANDING OUR HORIZONS IN VERIFICATION, VALIDATION, AND ACCREDITATION RESEARCH AND PRACTICE

Osman Balci
Richard E. Nance
James D. Arthur

Department of Computer Science
Virginia Tech
Blacksburg, VA 24061, U.S.A.

William F. Ormsby

Naval Surface Warfare Center
Dahlgren Division, Code T10
17320 Dahlgren Road
Dahlgren, VA 22448, U.S.A.

ABSTRACT

Many different types of modeling and simulation (M&S) applications, consisting of a combination of software, hardware, and humanware, are used in dozens of disciplines under diverse objectives including acquisition, analysis, education, entertainment, research, and training. Certification of sufficient accuracy of an M&S application by conducting verification, validation, and accreditation (VV&A) requires multifaceted knowledge and experience, and poses substantial technical and managerial challenges for researchers, practitioners, and managers. The challenges can only be met by using a very broad spectrum of approaches and expanding our horizons in VV&A. This paper presents 13 strategic directions to meet those challenges. The strategic directions provide guidelines for successful VV&A research and practice.

1 INTRODUCTION

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 Instruction 5000.61 (DoDI 1996) states that "It is the DoD policy that: … models and simulations used to support major DoD decision-making organizations and processes … shall be accredited for that use by the DoD component sponsoring the application."

Assuring the credibility of all those diverse types of complex M&S applications for acquisition, analysis or training poses significant technical and managerial challenges for researchers, practitioners, and managers. We must use a very broad spectrum of approaches and expand our horizons in VV&A in order to meet those challenges. In this paper, we propose 13 strategic directions to achieve this goal.

2 EXPANDING VV&A FROM ACCURACY-CENTERED ASSESSMENT TO QUALITY-CENTERED ASSESSMENT

Verification deals with transformational accuracy. *Validation* deals with behavioral or representational accuracy. *Accreditation* is defined as "the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose" (DoDI 1996). The term "acceptable" commonly refers to accuracy; accreditation therefore is conducted mainly for accuracy assessment.

One of the principles of testing dictates that complete M&S testing is not possible (Balci 1997). Exhaustive (complete) testing requires testing the M&S application under all possible inputs. Combinations of feasible values of M&S input variables can generate millions of logical paths during execution. Due to time and budgetary constraints, testing the accuracy of so many logical paths is impossible. Therefore, we say that "the only exhaustive testing there is, is so much testing that the tester is exhausted." Consequently, in M&S testing, the purpose is to increase our confidence in M&S accuracy as much as is dictated by the M&S intended uses and project objectives rather than trying to test the M&S application completely.

Although more than 100 verification and validation (V&V) techniques (Balci 1998; Binder 2000) are available, only a limited number of techniques are used in an M&S project due to time and resource constraints. Limited testing hinders our ability to substantiate sufficient M&S accuracy. Muessig, Laack, and Wrobleski (2000) indicate that validation alone has well-known limitations and advocate the assessment of accuracy together with M&S quality characteristics.

In those cases where the M&S application represents a non-existent system or in many other cases, we rely on the knowledge of subject matter experts (SMEs) in judging M&S accuracy. However, identifying and employing high quality SMEs is known to be very difficult (Glasow 1998; Pace 1998). Therefore, due to subjective human judgment and imperfect SME knowledge, no concrete conclusion can be drawn about the accuracy of a reasonably complex M&S application with 100% assurance.

Under the current V&V technology, we are unable to claim sufficient accuracy of a reasonably large and complex M&S application with 100% confidence due to M&S complexity, reliance on qualitative human judgment, and lack of complete testing. Sargent (1999) points out that "Instead, tests and evaluations are conducted until sufficient confidence is obtained that a model can be considered valid for its intended application." Therefore, the general consensus is that M&S V&V should be viewed as a "confidence building" activity.

For a reasonably large and complex M&S application, the "confidence building" activity must be performed by considering not only the M&S accuracy, but also other quality characteristics that affect acceptability, and hence, accuracy. The software quality characteristics are commonly called "ilities" in software engineering and include the following: adaptability, availability, complexity, fault-tolerance, maintainability, observability, performance, portability, readability, reliability, reusability, safety, security, survivability, testability, traceability, and usability.

Assuring a quality characteristic may enable us to increase our confidence in the M&S accuracy. For example, if the M&S application is assured to possess high testability, we may have higher confidence in M&S accuracy as opposed to the case where the M&S application is developed in a tightly integrated manner with poor testability. Conversely, poor traceability may increase the complexity of testing and reduce our confidence. Other quality characteristics such as readability, usability, observability, maintainability, reusability, and adaptability might also serve to increase our confidence in M&S accuracy.

In software engineering, software quality assurance (SQA) is a practice for assuring the required quality characteristics of a software product including accuracy (Schulmeyer and McManus 1999). An M&S application developed under an effective SQA program increases our confidence in accuracy much more than the M&S application developed without an SQA program.

In software engineering and other disciplines, accreditation is referred to as "certification" and is carried out for the quality characteristics. The following references describe how certification is currently practiced in software engineering: (ISACC 1999, 2000; Loesh et al. 1999; Poore, Mills, and Mutchler 1993; Rae, Robert, and Hausen 1995; Rodríguez-Dapena 1999; Vermesan 1997, 1998; Voas 1998a,b,c, 1999a,b,c,d, 2000a,b; Wakid, Kuhn, and Wallace 1999; Wallace 1999; Wohlin and Regnell 1998; Wohlin and Runeson 1994).

Using the software engineering terminology, M&S accreditation corresponds to M&S certification for the accuracy quality characteristic. However, if the M&S application is a physical product (e.g., a flight simulator used for training) that consists of hardware and software, quality characteristics such as maintainability, usability, reliability, safety, expandability, and adaptability may also be required in addition to the accuracy quality characteristic. In this case, it may be necessary to certify the M&S application for all those required quality characteristics, and not only for accuracy.

Currently, the term "accreditation" is interpreted in a narrowly defined manner. Such narrow-minded interpretation hinders our ability to do a better job in certifying M&S applications. We should expand our horizons by studying how software engineering and other disciplines assess accuracy within the context of certification and total quality management (TQM).

3 EXPANDING VV&A FROM PRODUCT-CENTERED ASSESSMENT TO (PRODUCT / PROCESS / PROJECT)-CENTERED ASSESSMENT

In software engineering, measurement and evaluation are commonly carried out in terms of product, process, and project (Rae, Robert, and Hausen 1995; Arthur and Henry 1995; Arthur, Nance, and Balci 1993; Freedman and Weinberg 1990). "Products" refer to the artifacts, created during the development life cycle, such as the requirements specification document, design specification document, or code. "Process" refers to a series of activities conducted to create a life-cycle product. Example processes include requirements engineering, design, and code production. "Project" deals with the management issues including personnel, resources, planning, and control.

Much of the M&S VV&A research and practice have concentrated on product measurement and evaluation. Because of the reasons mentioned in the previous section, under the current VV&A technology, we are unable to claim sufficient accuracy of an M&S life cycle product with 100% confidence. Therefore, we must examine the processes employed in creating the M&S application as well as how the M&S development activity is managed so as to increase our confidence in M&S accuracy.

Carnegie Mellon University Software Engineering Institute (CMU SEI 1994) has developed the Capability Maturity Model (CMM) as an application of the process management concepts of TQM to software. CMM is now very commonly used in the software industry as a means of judging software development process maturity and quality.

CMM can certainly be applied to M&S application development as well. Pointing out that "… the cost and effort

involved with VV&A will be significantly reduced once good software engineering discipline is enforced through acquisition practices", Conwell, Enright, and Stutzman (2000, pp. 827-828) introduce CMM in support of M&S VV&A. An M&S application developed by an organization at CMM Level 3 is expected to carry more credibility than if it were developed by an organization at Level 1.

For software certification, Voas (1998c, 1999b) proposes the software quality certification triangle, which consists of certification from three perspectives: product, process, and personnel. He advocates that we can approach certification from any one of the three perspectives, but a combination of all three will provide the best balance.

4 DEVELOPING M&S APPLICATIONS USING THE COMPONENT-BASED TECHNOLOGY

A recent symposium focused on making VV&A effective and affordable (Glasow and Pace 1999). It was noted at the symposium that component-based development is indeed an effective and affordable way of creating M&S applications and conducting M&S VV&A. A verified and validated model component can substantially decrease the M&S VV&A effort when reused. Such practice can significantly decrease the time and cost of M&S development. Component-based M&S development technology might be the "silver bullet" for effective and affordable M&S VV&A.

Component-based development is becoming increasingly important (Brown 1996). Component-based software development technology creation is currently led by NIST under the advanced technology focused program on Component-Based Software (NIST 2002). NIST cites many advantages of component-based development that can be realized for M&S application development conditioned on the following:

1. Establishment of a marketplace for component-based M&S application development so that M&S technology users can realize significant economic benefits through (a) reduced M&S project costs, (b) enhanced M&S credibility, and (c) expanded applicability of less expensive technology.
2. Increased automation and productivity in M&S application development enabling (a) improved M&S quality characteristics, (b) reduced time to develop, test, and certify M&S applications, and (c) increased amortization of costs through M&S component reuse.
3. Increased productivity of M&S project teams by (a) permitting specialists in the application domain to create higher fidelity M&S components, and (b) providing a focus on discourse in M&S development at a level far more comfortable to application domain users than a programming language.
4. Expanded markets for M&S application and component producers by promoting (a) the creation of systematically reusable M&S components, (b) increased interoperability among M&S software and non-M&S software products, and (c) convenient and ready adaptation of M&S components.

Voas (1998a) indicates that "Off-the-shelf components [COTS] could save the software industry considerable time and money. However, the industry first needs a set of black-box processes to certify the suitability of COTS components." Wohlin and Regnell (1998) present an approach for reliability certification of software components.

5 PRACTICING VV&A IN AN INDEPENDENT MANNER

Independent verification and validation (IV&V) is a technique of long-standing in the field of software engineering. The IEEE Standard for V&V states that "the IV&V responsibility is vested in an organization that is separate from the development organization" (IEEE 1998, p. 58). Boehm (1984, p. 76) advocates that "verification and validation activities produce their best results when performed by a V&V agent who operates independently of the developer or specification agent." IV&V is a cost-effective way to mitigate the many risks inherent in large-scale software development projects (IEEE 1998; Arthur et al. 1999; Arthur and Nance 2000; Lewis 1992).

VV&A should be conducted under technical, managerial, and financial independence (IEEE 1998). Technical Independence implies that the VV&A agent determines, prioritizes, and schedules its own tasks and efforts. Managerial Independence implies that the VV&A agent reports to the M&S application sponsor independently of the developer organization. Financial Independence implies that the VV&A agent is allocated its own budget for the M&S VV&A and does not rely on the M&S development budget.

Several scenarios exist under which independence can be achieved at different levels as depicted in Figure 1. The M&S application developer can promote IV&V within its organizational structure under three strategies:

1. Establish an SQA group, which reports to the upper management independently from the M&S development group.
2. Subcontract the V&V work to another organization (V&V agent), which reports to the upper management independently from the M&S development group.
3. Perform both (1) and (2) above.

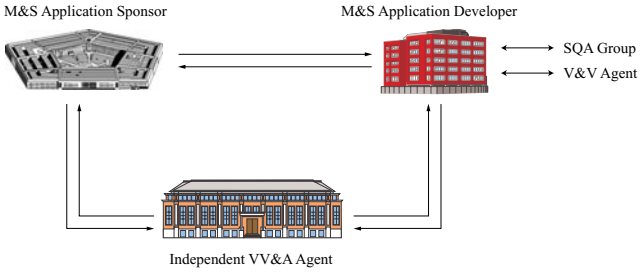

Figure 1: Achieving Independence at Different Levels

True independence can only be achieved when the M&S application sponsor employs a VV&A agent, which is completely independent from the developer organization. The VV&A agent must be able to perform its responsibility under technical, managerial, and financial independence. It independently conducts M&S acceptability assessment and formulates a recommendation for accreditation / certification.

In those cases where the M&S application is already developed or is a legacy M&S application, the VV&A agent is called the accreditation agent. If the M&S application is under development, the VV&A agent should be hired as soon as the development is initiated. Two strategies exist in this case: (a) the agent is hired as an independent accreditation agent, who only performs the acceptability assessment at the end of the development life cycle, or (b) the agent is hired as an independent VV&A agent, who participates in the V&V activities and performs acceptability assessment throughout the entire development life cycle. The latter strategy offers many advantages, including:

- The developer gets feedback for acceptability as the M&S application development progresses throughout the life cycle,
- M&S application errors and deficiencies are discovered early in the life cycle,
- The complexity of M&S application acceptability assessment is reduced,
- Communication between the independent VV&A agent and the developer helps improve the M&S application quality,
- The developer is encouraged to implement an effective SQA program, and
- M&S application product and resource risks are significantly reduced.

6 PROVIDING COMPUTER-AIDED SUPPORT FOR VV&A

Osterweil et al. (1996, p. 738) state that "a number of studies have suggested that 50–60% of the effort involved in producing large software systems is devoted to quality assessment activities such as testing, and the percentage may be significantly higher for life-critical software systems." They indicate that to meet substantial software quality challenges, "it is necessary to improve the current tools, technologies, and their cost-benefit characterizations."

Each build or version of an M&S application, throughout its development life cycle, is subjected to testing so as to perform verification, validation, and/or acceptability assessment (to reach an accreditation decision). Such testing is designed by identifying test scenarios, test cases, and/or test data. Applying the same testing to each M&S build or version is certainly repetitive and time consuming, and requires the preservation of the test scenarios, test cases, and test data for reuse. Effective software tools providing computer-aided support for M&S VV&A can significantly reduce the testing time and effort.

A simulation model is built to represent a system with respect to the modeler's intent. SMEs are commonly used to compare model execution behavior to intent in conducting VV&A. To a certain extent, however, computer-aided support can be provided to perform this comparison by using the assertion checking V&V technique (Balci 1998). An *assertion* is a statement that should hold true as the simulation model executes. Assertion checking is a V&V technique used to compare an execution profile against the expectations of the modeler, and hence, guards model execution against potential errors. Software tools for M&S application development must facilitate the design and implementation of assertion checking.

More than 200 utilities and tools are available in the commercial market providing computer-aided support for software testing (Cigital 2002). Certainly, these tools can also be used for M&S VV&A. However, software tools specifically created for M&S VV&A (e.g., tools for experimental design V&V, random number generator V&V, random variate generator V&V) would provide much more effective computer-aided support.

7 DEVELOPING EFFECTIVE TECHNIQUES FOR MEASUREMENT AND EVALUATION OF QUALITATIVE ELEMENTS

Much of the assessment underlying M&S VV&A requires the measurement and evaluation of qualitative elements such as conceptual model credibility, data credibility, degree of model representativeness, M&S design process credibility, M&S integration risk, M&S requirements credibility, model maintainability, model validity, model verity, and project management quality. Measurement and evaluation of such qualitative elements pose substantial technical challenges.

In some cases, we may not be able to assess the sufficient credibility of the representation of some qualitative elements such as "threat" in M&S applications due to the lack of reliable information and inadequate SME knowledge. New techniques are needed to characterize the overall effect of such a "missing link" on the accreditation decision.

Balci (2001) presents an indicator-based certification / accreditation methodology that incorporates SME knowledge and measurement and evaluation of qualitative elements. Balci et al. (2002) present a collaborative Web-based Evaluation Environment software system (Orca 2002) for facilitating the application of the certification / accreditation methodology. Oberkampf et al. (2000) present a general methodology for estimating total uncertainty in computational simulations.

8 DEVELOPING EFFECTIVE TECHNIQUES FOR UTILIZING SME KNOWLEDGE

We face major challenges in eliciting, representing, and integrating the knowledge of SMEs employed for the VV&A tasks. Usually, a large volume of documentation (e.g., a requirements specification might extend to five volumes) is given to an SME to conduct a VV&A task (e.g., requirements V&V). Due to the complexity of the subject and large amount of information to comprehend, SMEs commonly provide unstructured, vague, and incomplete evaluations. Generally, one SME is not sufficiently knowledgeable about every aspect of the subject. Therefore, a number of SMEs are employed and each is assigned to the assessment of certain aspects of the subject matter. Then, the integration of SME assessments to reach an overall VV&A decision becomes a serious problem.

Effective techniques are needed for SME knowledge elicitation, representation, and integration. Ford and Sterman (1997) indicate that "increased clarity and specificity are required concerning the methods used to elicit expert knowledge for modeling." Although their work deals with knowledge elicitation to improve model development, the same or similar approaches can also be used for VV&A. Rush and Wallace (1997) state that "this process proves most difficult when the elicitation and representation of knowledge from multiple experts is necessary." Balci (2001) presents an indicator-based methodology for structuring and integrating SME evaluations. Ketcham and Muessig (2000) describe the need for, and benefit of, proper integration of SMEs into M&S design and implementation, and especially into M&S requirements management.

9 PRACTICING ACCREDITATION AND CERTIFICATION UNDER A COMPREHENSIVE SCHEME

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. We use the ISO terminology in this section.

Similar to the manner accreditation and certification are carried out in other disciplines, we propose the comprehensive scheme shown in Figure 2. The scheme assumes two scenarios: (a) M&S application development under contract, and (b) the component-based M&S development marketplace where developers fabricate reusable M&S components for sale.

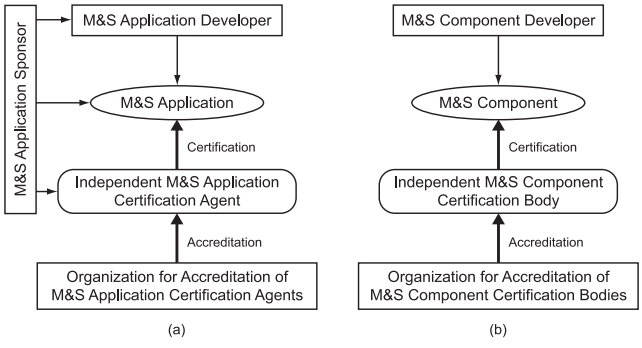

Figure 2: Accreditation and Certification Practice

Under scenario (a), an M&S application sponsor hires an independent M&S application certification agent, which is accredited by an external accreditation authority to conduct the required certification. Currently, this scenario is practiced in DoD under the label of "accreditation", where no authority exists to accredit the work of the agent. Certainly, the M&S application sponsor should be assured that the practice of the certification agent satisfies minimum standards established by an independent accreditation authority yet to be founded.

Under scenario (b), an M&S component developer fabricates a reusable M&S component for sale. The buyer of such a component seeks assurance about the component quality. An independent M&S component certification body can provide such an assurance by awarding a "mark of conformity" (e.g., certified to be HLA compliant), a "seal of approval", or a "certificate of excellence." There may be many certification bodies, which should provide certifications that are unbiased, fair, cost-effective, and reproducible. Therefore, the certification bodies should be accredited by an external accreditation authority. For example, the SGS International Certification Services (SGS-ICS) group of companies (http://www.sgsgroup.com/) is a leading international certification body, which is accredited in the United Kingdom by the National Accreditation Council for Certification Bodies (NACCB) and in Belgium by the Dutch Council for Certification (RvC).

In the United States, for example, the Accreditation Board for Engineering and Technology (ABET) accredits educational programs that award diplomas and certificates. ABET (2002) states that "The diversity of educational programs in the United States is one strength of the American educational system. Such a large selection of educational offerings makes quality a vital issue. Accreditation is the quality assurance that education is meeting minimum standards."

When the component-based M&S development technology is commonly used and an M&S component marketplace is established, organizations such as SGS-ICS can be founded to independently perform M&S component certification. Organizations similar to NACCB, RvC, and ABET can be founded to accredit, regulate, and monitor the M&S component certification bodies.

Legal issues about the practice of M&S certification and accreditation pose other challenges. The contractual agreement between the M&S application sponsor and developer should clearly specify the legal liability in the case of the M&S application failing the certification. What happens when DoD acquires an inadequate or unusable system, costing taxpayers millions of dollars, based on the use of an M&S application certified by an accredited agent? Who would be legally liable for the costly wrong decision? Rae, Robert, and Hausen (1995) discuss the legal issues in detail for software certification.

A certification program should also be instituted for the M&S quality assurance (QA) (i.e., VV&A) professionals. Similar to certification of engineers, providing certification for M&S QA professionals advances the M&S QA practice from ad hoc to a disciplined level. Because university or college degrees are not offered specifically for M&S QA, such a certification program is required to issue a license for M&S QA professionals. Currently, the American Society for Quality (http://www.asq.org/) provides such certifications for quality engineers and professionals.

Many organizations provide technical services for software product, process, and personnel certification including: American Society for Quality, Software Certification (http://www.isci.com/), Independent Quality Assurance Institute (http://www.qaiusa.com/), and Det Norske Veritas (http://www.dnv.com/certification), and TUV (http://www.tuvps.com/).

Many standards and guidelines are available for software evaluation and certification (Rae, Robert, and Hausen 1995). Table 1 lists some of them.

10 RESOLVING THE TERMINOLOGY PROBLEM

Many different types of M&S exist including combined simulation, continuous simulation, discrete-event simulation, distributed simulation, gaming, hardware-in-the-loop simulation, human-in-the-loop simulation, Monte Carlo simulation, parallel simulation, synthetic environments bringing together simulations and real-world systems, system dynamics simulation, and system theoretical simulation. The use of these different M&S types spans dozens of different disciplines for many different purposes including acquisition, analysis, education, entertainment, research, and training.

Table 1: Standards and Guidelines Related to Software Evaluation and Certification

Generic Evaluation and Certification	
ISO/IEC Guide 25	General Requirements for the Technical Competence of Testing Laboratories
ISO/IEC Guide 28	General Rules for a Model Third-Party Certification System for Products
ISO/IEC Guide 40	General Requirements for the Acceptance of Certification Bodies
Process Evaluation and Certification	
ISO 9000	Quality Management and Quality Assurance Standards
IEEE 1298	Software Quality Management System
ISO 9001	Quality Systems – Model for Quality Assurance in Design/Development, Production, Installation and Servicing
Product Evaluation and Certification	
IEEE 1061	Standard for a Software Quality Metrics Methodology
ISO/IEC 9126	Software Product Evaluation
ISO/IEC DIS 12119	Quality Requirements and Testing
Quality Assurance	
ANSI/IEEE 730	Software Quality Assurance Plans
AQAP-1	NATO Requirements for an Industrial Quality Control System
DoD STD 2168	Defense System Software Quality Program

Such broad and diverse use of M&S across many disciplines creates a serious terminology problem. Even the commonly accepted definitions for VV&A terms are interpreted in such different ways that misunderstandings and communication confusion abound in the VV&A practice and research (Glasow and Pace 1999). Glasow and Pace (1999) report that "the breadth and depth of potential trouble that such communication difficulties can cause the M&S VV&A community are not fully appreciated by either VV&A practitioners or M&S management."

Probably the most serious terminology problem is the definition and use of the term "accreditation" in the M&S community. It conflicts with the definition consistently adopted by ISO, software engineering, and other disci-

plines. DoD Instructions should be changed to use the term "certification" to be consistent with the other disciplines.

11 IMPROVING THE QUALITY OF THE VV&A PLAN AND ITS EXECUTION

The VV&A plan is developed by the independent VV&A agent as the blueprint of VV&A activities throughout the development life cycle (Balci et al. 2000). The quality of the plan and the quality of its execution affect the confidence level at which the accreditation decision is made as depicted in Figure 3. The impact of the quality on the confidence level is situation-dependent and changes from one M&S project to another as shown by the different curves with shape parameters α_i, i=1,2,3,4 in Figure 3.

Comprehensiveness is one of the most important quality characteristics of a VV&A plan. By analogy, if an M&S application corresponds to a "forest", its VV&A plan describes how the "forest" will be evaluated. If a VV&A plan is structured to evaluate only, for example, 30% of the "trees" and their "branches", the accreditation decision can only be based on 30% of the "full picture" resulting in a low confidence level.

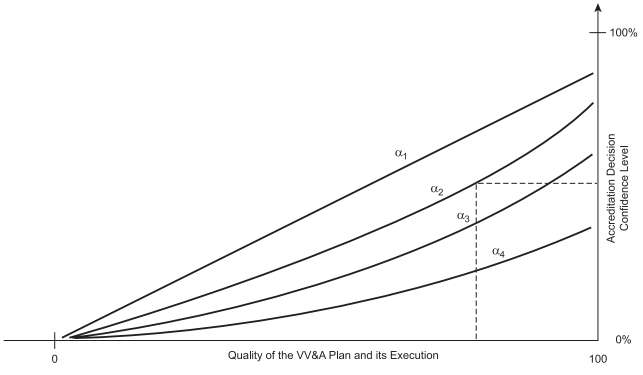

Figure 3: VV&A Plan and Execution Quality versus Accreditation Decision Confidence Level

Sometimes, it is not possible to execute a VV&A plan completely due to factors such as lack of resources, changing requirements, and development refocus. In this case, the plan should show how much VV&A is conducted with respect to the comprehensive set of requirements specified in the plan. The amount of coverage should be taken into consideration in determining the confidence level at which the accreditation decision is recommended to the M&S application sponsor by the VV&A agent. Under no circumstance, should the VV&A agent develop and use an incomplete VV&A plan. If the VV&A agent develops and uses an incomplete plan and if the M&S application fails the accreditation, the M&S developer may direct blame to the VV&A agent and attribute the failure to the incomplete VV&A plan.

Having a sufficiently credible VV&A plan is a necessary, but not a sufficient condition. The quality of the execution of the VV&A plan is also very important. Successful execution of the plan requires an effective SQA program established by the M&S developer and an effective cooperation between the M&S developer and the VV&A agent. Quality of the SMEs used for the VV&A activities significantly affects the quality of execution of the plan.

Planning for the VV&A activities should be coordinated among the M&S application sponsor, M&S developer's SQA group, V&V agent, and the independent VV&A agent. A VV&A plan should generally describe:

- organizational responsibilities,
- well-defined intended uses of the M&S application (Balci and Ormsby 2000),
- acceptability criteria (also called accreditation criteria),
- the VV&A methodology or structured approach to be used,
- indicators for acceptability assessment for each phase of the development life cycle, and
- a schedule (if possible).

12 IMPROVING THE RELATIONSHIP BETWEEN THE M&S APPLICATION DEVELOPER AND THE INDEPENDENT VV&A AGENT

Successful independent VV&A requires the VV&A 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 VV&A agent. Sometimes, developers view independent VV&A as a performance appraisal activity, and they fear that their reputation and potential future funding are at stake if the VV&A agent identifies problems. Therefore, they sometimes show no desire to cooperate and behave in an adversarial manner against the independent VV&A personnel.

The M&S application sponsor has a critical role in resolving this problem. The sponsor must clearly explain to the developer that the independent VV&A activity is not intended to be a performance appraisal activity, and the overall goal for all parties involved is to produce an acceptable M&S application. The independent VV&A agent is hired to work together with the developer to accomplish that goal. Independent VV&A cannot be successfully conducted without productive cooperation between the developer and the VV&A agent.

The sponsor must explicitly specify the requirements for the working relationship between the developer and the VV&A agent in the legal contractual agreement signed between the sponsor and the developer. Without legal con-

tractual requirements, the developers are often less motivated to cooperate with the independent VV&A agents. The M&S application sponsor must make sure that the legal contract is written in such a way that the productive cooperation between the developer and the VV&A agent is legally secured.

13 PROVIDING COMPREHENSIVE EDUCATION FOR VV&A

The education required for conducting successful VV&A is multifaceted and cannot be provided within a single university-level curriculum. A person is commonly educated in some of the facets but not in all. Generally, two types of education are required: technical and managerial. Although technical education is mostly emphasized, managerial knowledge is at least as important as technical.

Many people influence the success of VV&A by playing different roles and requiring different types of knowledge (see Figure 1):

- M&S application sponsor's managers,
- M&S developer's project managers,
- M&S developer's SQA engineers and managers,
- the V&V agent's engineers and managers, and
- the independent VV&A or accreditation agent's engineers and managers.

The M&S application sponsor's managers should be well educated about VV&A, especially in areas such as the importance of VV&A as a means of M&S project risk mitigation, benefit-cost analysis, additional resources required for the developer to establish an effective SQA program, and the requirement for an independent VV&A agent. If the sponsor's managers fail to provide adequate VV&A funding, they cannot expect a quality product or a successful M&S project.

The M&S developer's project managers should be well educated in project management, including areas such as application development management, asset and technical infrastructure management, configuration management, management metrics and measurement, managing technical change and legacy systems, quality management, requirements and test management, risk management, software lifecycle economics, software process definition and improvement, and team management and development.

The M&S developer's SQA engineers and managers should be well educated in SQA, which is a profession, in and of itself, and having its own handbook (Schulmeyer and McManus 1999).

The V&V agent's engineers and managers should be well educated in the V&V technology as well as in SQA.

The independent VV&A or accreditation agent's engineers and managers should be well educated in the V&V technology, test and evaluation, accreditation/certification, and SQA.

14 DISSEMINATING THE VV&A EXPERIENCE

Although much VV&A work is carried out, and especially in DoD, very little work is published describing the VV&A experience. The classified or confidential nature of the DoD VV&A work precludes the publication of papers in the open literature (e.g., conference proceedings and journals). However, it is often possible to create an unclassified description of the work and write a paper for disseminating the VV&A experience. A paper can be written about the VV&A experience without identifying the M&S application and the organizations involved.

In those cases where a VV&A paper can be written, contractors are unwilling to develop the paper and present it at a conference due to lack of funding. Usually, a DoD contract does not provide funding for attending a conference to present a paper, and contractors do not want to pay for the expenses from the company overhead budget due to lack of incentives. Today, with manpower shortages, the incentive for attending conferences and publishing papers for professional visibility might be necessary to retain exceptional technical talent.

Failing to disseminate the VV&A experience may result in the repetition of the same mistakes in future M&S projects. Based on past experience, establishing a better quality assurance program for a new M&S project may certainly increase the probability of success for that project. Learning from the past experience of others is an excellent and cost-effective educational tool.

DoD sponsors should provide funding and incentives for the contractors to publish their VV&A work. The return on such an investment can easily be realized by preventing the failures of M&S projects or by preventing wrong simulation-based acquisition decisions.

15 CONCLUDING REMARKS

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 TQM programs, and software development companies have SQA departments to be able to meet the quality challenge.

M&S applications are 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. Software version 1.0 commonly implies "we hope it works!" There is no "silver bullet". A disciplined approach, effective management, and well-educated personnel are some of the key factors affecting the success of a software development project.

We strongly advocate the position that M&S professionals can learn a lot from software engineering and many other disciplines. In so doing, we can expand our horizons to include necessary knowledge for conducting successful M&S VV&A.

Although DoD Instruction 5000.61 (DoDI 1996) requires the VV&A of models and simulations used to support major DoD decision-making processes, the DoD and other government funding agencies do not provide significant support for VV&A basic research. Glasow in her position statement (Sargent et al. 2000, p. 909) indicates that little research has been undertaken within the DoD and contends that "the focus has been on establishing a baseline for VV&A practice across DoD, in the form of policies and procedures, rather than the conduct of scientifically rigorous research." Significant funding is critically needed for VV&A basic research to be able to meet the technical challenges we currently face.

ACKNOWLEDGMENT

The research leading to the development of this paper has been funded by the Naval Surface Warfare Center Dahlgren Division, Dahlgren, Virginia.

REFERENCES

ABET. 2002. Accreditation board for engineering and technology. Baltimore, MD. http://www.abet.org

Arthur, J. D., M. K. Groener, K. J. Hayhurst, and C. M. Holoway. 1999. Evaluating the effectiveness of independent verification and validation. *IEEE Computer* 32 (10): 79-83.

Arthur, J. D. and S. M. Henry. eds. 1995. Software process and product measurement. *Annals of Software Engineering*. Vol. 1.

Arthur, J. D. and R. E. Nance. 2000. Verification and validation without independence: a recipe for failure. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 859-865. Piscataway, NJ: IEEE.

Arthur, J. D., R. E. Nance, and O. Balci. 1993. Establishing software development process control: technical objectives, operational requirements, and the foundational framework. *Journal of Systems and Software* 22 (2): 117-128.

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. 2002. 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. 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., W. F. Ormsby, J. T. Carr III, and S. D. Saadi. 2000. Planning for verification, validation, and 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, 829-839. Piscataway, NJ: IEEE.

Binder, R. V. 2000. *Testing Object-Oriented Systems: Models, Patterns, and Tools*. Reading, MA: Addison-Wesley.

Boehm, B. W. 1984. Verifying and validating software requirements and design specifications. *IEEE Software*. 75-88.

Brown, A. W. ed. 1996. *Component-Based Software Engineering*. Los Alamitos, CA: IEEE CS Press.

Cigital. 2002. Testing tools supplier list. Cigital, Dulles, VA. http://www.testingfaqs.org/tools.htm

CMU SEI. 1994. *The Capability Maturity Model: Guidelines for Improving the Software Process*. Reading, MA: Addison-Wesley.

Conwell, C. L., R. Enright, and M. A. Stutzman. 2000. Capability maturity models support of modeling and simulation verification, validation, and accreditation. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 819-828. Piscataway, NJ: IEEE.

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

Ford, D. N. and J. D. Sterman. 1997. Expert knowledge elicitation to improve mental and formal models. *System Dynamics Review* 14 (4): 309-340.

Freedman, D. P. and G. M. Weinberg. 1990. *Handbook of Walkthroughs, Inspections, and Technical Reviews:*

Evaluating Programs, Projects, and Products. New York, NY: Dorset House Publishing.

Glasow, P. A. 1998. Characteristics, selection, and use of subject matter experts. Tech. Report 99W0000048, POET WG2K IV&V Paper No. 25. McLean, VA: The MITRE Corporation.

Glasow, P. A. and D. K. Pace. eds. 1999. Report on the simulation validation—making VV&A effective and affordable—Mini-Symposium and Workshop. Alexandria, VA: Military Operations Research Society.

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

ISACC. 1999. Proc. of the 1st international software assurance certification conference. http://www.isacc.com/

ISACC. 2000. International software assurance certification conference. http://www.isacc.com/

Ketcham, R. L. and P. R. Muessig. 2000. Software engineering and simulation credibility. In *Proceedings of the 2000 Summer Computer Simulation Conference*. San Diego, CA: SCS.

Lewis, R. O. 1992. *Independent Verification & Validation: A Life Cycle Engineering Process for Quality Software*. New York, NY: John Wiley & Sons.

Loesh, R. E., A. B. Gosnell, T. F. Benoist, R. M. Wyskida, and J. D. Johannes. 1999. An engineering approach to critical software certification. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*. Washington, DC: IEEE Computer Society.

Muessig, P. R., D. R. Laack, and J. J. Wrobleski. 2000. An integrated approach to evaluating simulation credibility. In *Proceedings of the 2000 Summer Computer Simulation Conference*. San Diego, CA: SCS.

NIST. 2002. Component-based software advanced technology program. NIST, Gaithersburg, MD. http://www.atp.nist.gov/atp/focus/cbs.htm

Oberkampf, W. L., S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin. 2000. Estimation of total uncertainty in modeling and simulation. Report SAND2000-0824, Sandia National Laboratories, Albuquerque, NM, Apr.

Orca. 2002. Evaluation environment. Orca Computer, Inc., Blacksburg, VA, https://www.orcacomputer.com/ee

Osterweil, L. et al. 1996. Strategic directions in software quality. *ACM Computing Surveys* 28 (4): 738-750.

Pace, D. K. 1998. Verification, validation, and accreditation. In *Applied Modeling and Simulation: An Integrated Approach to Development and Operation*. ed. D. J. Cloud and L. B. Rainey, 369-410. New York, NY: McGraw-Hill.

Poore, J. H., H. D. Mills, and D. Mutchler. 1993. Planning and certifying software system reliability. *IEEE Software*. 88-99.

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

Rodríguez-Dapena, P. 1999. Software safety certification: a multidomain problem. *IEEE Software* 16 (4): 31-38.

Rush, R. and W. A. Wallace. 1997. Elicitation of knowledge from multiple experts using network inference. *IEEE Transactions on Knowledge and Data Engineering* 9 (5): 688-696.

Sargent, R. G. 1999. Validation and verification of simulation models. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 39-48. Piscataway, NJ: IEEE.

Sargent, R. G., P. A. Glasow, J. P. C. Kleijnen, A. M. Law, I. McGregor, and S. Youngblood. 2000. Panel: strategic directions in verification, validation, and accreditation research. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 909-916. Piscataway, NJ: IEEE.

Schulmeyer, G. G. and J. I. McManus. eds. 1999. *Handbook of Software Quality Assurance*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Vermesan, A. I. 1997. Quality assessment of knowledge-based software: some certification considerations. In *Proceedings of the 3rd International Software Engineering Standards Symposium* (June 1-6, Walnut Creek, CA), IEEE Computer Society, Washington, DC, pp. 144-154.

Vermesan, A. I. 1998. Software certification for industry – verification and validation issues in expert systems. In *Proceedings of the 9th International Workshop on Database and Expert Systems Applications*. Washington, DC: IEEE Computer Society.

Voas, J. M. 1998a. Certifying off-the-shelf software components. *IEEE Computer* 31 (6): 53-59.

Voas, J. M. 1998b. Certifying high assurance software. In *Proceedings of the 22nd International Computer Software and Application Conference*. Washington, DC: IEEE Computer Society

Voas, J. M. 1998c. The software quality certification triangle. *Crosstalk* 11 (11): 12-14.

Voas, J. M., ed. 1999a. Special issue on software certification. *IEEE Software* 16 (4).

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

Voas, J. M. 1999c. Certifying software for high-assurance environments. *IEEE Software* 16 (4): 48-54.

Voas, J. M. 1999d. Can generic software be assured? In *Proceedings of the 23rd Annual International Computer Software and Applications Conference.* Washington, DC: IEEE Computer Society.

Voas, J. M. 2000a. Limited software warranties. In *Proceedings of the 7th IEEE International Conference*

and Workshop on the Engineering of Computer Based Systems. Washington, DC: IEEE Computer Society.

Voas, J. M. 2000b. Developing a usage-based software certification process. *IEEE Computer* 33 (8): 32-37.

Wakid, S. A., D. R. Kuhn, and D. R. Wallace. 1999. Toward credible IT testing and certification. *IEEE Software* 16 (4): 39-47.

Wallace, D. R. 1999. Software verification & validation's role in product certification. In *Proceedings of the First International Software Assurance Certification Conference*. Cigital, Sterling, VA.

Wohlin, C. and B. Regnell. 1998. Reliability certification of software components. In *Proceedings of the 5th International Conference on Software Reuse*. Washington, DC: IEEE Computer Society.

Wohlin, C. and P. Runeson. 1994. Certification of software components. *IEEE Transactions on Software Engineering* 20 (6): 494-499.

AUTHOR BIOGRAPHIES

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 is the Editor-in-Chief of the international journal *Annals of Software Engineering*; Verification, Validation and Accreditation (VV&A) Area Editor of *ACM Transactions on Modeling and Computer Simulation*; and Modeling and Simulation (M&S) Category Editor of *ACM Computing Reviews*. 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 DoD since 1983. Since 1998, he has been providing technical services for the National Missile Defense program. His current areas of expertise center on Software Engineering; E-Systems Engineering (e.g., 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>.

RICHARD E. NANCE is the RADM John Adolphus Dahlgren Professor of Computer Science and the Director of the Systems Research Center at Virginia Tech. Dr. Nance is also Chairman of the Board of Orca Computer, Inc. He has held a number of editorial positions and was the founding Editor-in-Chief of the *ACM Transactions on Modeling and Computer Simulation*, 1990-1995. Currently, he is a member of the Editorial Board, Software Practitioner Series, Springer. He served as Program Chair for the 1990 Winter Simulation Conference. Dr. Nance received a Distinguished Service Award from the TIMS College on Simulation in 1987. In 1995 he was honored by an award for "Distinguished Service to SIGSIM and the Simulation Community" by the ACM SIG on Simulation. He was named ACM Fellow in 1996. His email and web addresses are <nance@vt.edu> and <http://www.cs.vt.edu/info/people/vitae/Nance.html>.

JAMES D. ARTHUR is an Associate Professor of Computer Science at Virginia Tech. He received B.S. and M.A. degrees in Mathematics from the University of North Carolina at Greensboro in 1972 and 1973, and M.S. and Ph.D. degrees in Computer Science from Purdue University in 1981 and 1983. His research interests include Software Engineering (Methods and Methodologies supporting Software Quality Assessment and IV&V Processes), Parallel Computation, and User Support Environments. He is the author of over 30 papers on software engineering, software quality assessment, IV&V, and user/machine interaction. He has served as: participating member of IEEE Working Group on Reference Models for V&V Methods; Chair of Education Panel for National Software Council Workshop; and Co-Guest Editor for *Annals of Software Engineering* special volume on Process and Product Quality Measurement. His email and web addresses are <arthur@vt.edu> and <http://vtopus.cs.vt.edu/~arthur>.

BILL ORMSBY is currently a Program Manager for the Naval Surface Warfare Center, Dahlgren Division (NSWCDD). Bill began his career at NSWCDD as an Aerospace Engineer in 1981, after receiving his M.S. degree in Engineering from Virginia Tech, and has specialized in simulation-based system acquisition since that time. His work in the Space and Geodesy Branch has spanned all phases of system acquisition, from concept definition through test and evaluation. He spent 10 years defining, designing, implementing and testing embedded software systems in support of the National Imaging and Mapping Agency. The systems he developed were subjected to a rigorous V&V process to ensure reliable performance when integrated with critical space systems. From 1990 to 1995, Bill managed the TRIDENT II Guidance System IV&V Program and applied the same rigor to systems developed by Draper Labs. From 1995 to 2000, he managed the National Missile Defense (NMD) IV&V Program. During this time, Bill served as Co-Chair of the NMD M&S VV&A Subgroup and consistently applied best practices to achieve program objectives. Bill has relied heavily on modeling and simulation (M&S) as a method for verifying system performance in anticipated operational environments. He has been involved, in some capacity, in dozens of M&S development initiatives. His email address is <OrmsbyWF@nswc.navy.mil>.