

## A COLLABORATIVE EVALUATION ENVIRONMENT FOR CREDIBILITY ASSESSMENT OF MODELING AND SIMULATION APPLICATIONS

Osman Balci  
Robin J. Adams  
David S. Myers  
Richard E. Nance

Orca Computer, Inc.  
Virginia Tech Corporate Research Center  
1800 Kraft Drive, Suite 111  
Blacksburg, VA 24060, U.S.A.

### ABSTRACT

Credibility assessment of modeling and simulation (M&S) applications is becoming increasingly more important as M&S applications are used more and more for complex system design evaluation, M&S-based acquisition, problem solving, military training, and critical decision making. M&S credibility assessment is a very complex process, involves the measurement and evaluation of hundreds of qualitative and quantitative elements, mandates subject matter expert evaluation, and requires the integration of disparate measurements and evaluations. Planning and managing such measurements and evaluations require a unifying methodology and should not be performed in an *ad hoc* manner. We have developed such a methodology, which can be used for credibility assessment of any kind of M&S application. To provide computer-aided assistance in applying our methodology and to enable collaborative evaluations by geographically dispersed people, we have developed a Web-based client/server software system. This tutorial introduces our software system, which can be used at <https://www.orcacomputer.com/ee> or <http://www.EvaluationEnvironment.com>.

### 1 INTRODUCTION

The Evaluation Environment (EE) is a Web-based client/server software system that works through the Secure Sockets Layer (SSL) with 128-bit encryption on the Internet. It enables geographically dispersed people to conduct complex evaluation projects in a collaborative manner. EE can be used for any kind of evaluation project including modeling and simulation (M&S) credibility assessment.

M&S credibility (or acceptability) assessment can be conducted under the guidance of the following:

- A multifaceted methodology for certification of M&S applications (Balci 2001),
- Principles of M&S verification, validation, and testing (Balci 1997, 1998; DMSO 2000),
- Standard processes for certification of M&S applications (Balci and Saadi 2002),
- DoD M&S verification, validation, and accreditation (VV&A) policies and instructions (DMSO 2000; DoDI 1996),
- IEEE standards for software verification and validation (IEEE 1998), and
- Software certification best practices (Balci et al. 2002; Rae, Robert, and Hausen 1995; Voas 1999).

We refer the reader to the literature cited above for background information about M&S credibility assessment.

The purpose of this tutorial paper is to teach how to use EE for assessing overall M&S credibility (or acceptability). Section 2 presents step-by-step instructions for conducting an EE project for M&S credibility assessment. Concluding remarks are given in section 3.

### 2 M&S CREDIBILITY ASSESSMENT USING EE

In this section, we provide step-by-step instructions for using EE to assess the overall credibility (or acceptability) of an M&S application.

#### 2.1 Step 1: Project Identification and Setup

M&S credibility assessment can be conducted either for an already developed M&S application or for a particular development life cycle stage of an M&S application under

development. Balci and Saadi (2002) propose standard evaluation processes throughout the M&S development life cycle. Each evaluation process is conducted to assess the credibility of the corresponding life cycle stage.

Identify if an EE project will be created for credibility assessment of an already developed M&S application or for a development life cycle stage of an M&S application under development.

Identify the EE Project Administrators who will have full permission to make changes to the project. Identify subject matter experts (SMEs) who have technical and/or operational domain knowledge and expertise about the problem domain and M&S engineering. Have all Administrators and SMEs register as EE users.

## 2.2 Step 2: Creation of a Hierarchy of Indicators

Have all Administrators and SMEs create a hierarchy of indicators by working together in a collaborative manner.

Figure 1 shows the top two levels of a hierarchy of indicators for overall credibility or acceptability assessment of an already developed M&S application. The hierarchy is created to represent the proposed standard evaluation processes described by Balci and Saadi (2002).

M&S acceptability (or overall credibility) is a complex qualitative concept and cannot be directly measured. Therefore, to measure it indirectly, we decompose it at level 1 into credibility assessments of M&S requirements, conceptual model, M&S design, executable M&S modules, integrated M&S application, M&S results, and presented results.

Each of the seven credibility indicators at level 1 is also a complex qualitative concept and cannot be directly measured. Therefore, to measure each indirectly, we decompose each further at level 2 into the quality indicators as shown in Figure 1.

Each quality indicator at level 2 is also a complex qualitative concept and cannot be directly measured. Therefore, to measure each indirectly, we should decom-

The screenshot displays a web browser window titled "EE: M&S Acceptability Assessment Indicator Browser - Microsoft Internet Explorer". The address bar shows "https://www.orcacomputer.com/ee/". The main content area is divided into two panes. The left pane shows a hierarchical tree of indicators under "M&S Acceptability". The tree is expanded to show "Conceptual Model Credibility", which includes "Conceptual Modeling Process Quality". The right pane shows the detailed view for "Conceptual Modeling Process Quality". It has tabs for "Description", "Weights", "Scores", "Evaluation", "Attributes", and "Evaluators". The "Description" tab is active, showing a "Definition:" section. The definition states: "According to DMSO [2000], the *Conceptual Modeling Process* involves the development of". Below the definition is a list of three bullet points:

- M&S context including functional description of the mission space, constraints, and bounds,
- M&S concept including mission space representation and M&S space functionality, and
- M&S elements including architectural relationships, assumptions, algorithms, and data.

Figure 1: Top Two Levels of a Hierarchy of Indicators for M&S Acceptability Assessment

pose each further at level 3. The decomposition should continue until the leaf indicators are assessable.

The hierarchy consists of three types of indicators: root, branch, and leaf.

- A *root indicator* is at the apex of the hierarchy, obviously having no parent. It represents the qualitative concept “M&S acceptability”.
- A *branch indicator* is one that has at least one parent indicator and at least one child indicator.
- A *leaf indicator* is one that has at least one parent indicator and no child indicator.

A leaf or branch indicator can be designated to influence more than one parent indicator using the Adopt button on the indicator browser toolbar. Therefore, the hierarchy represents a directed acyclic graph.

### 2.3 Step 3: Review of the Hierarchy of Indicators

Form an independent review panel consisting of SMEs, technical people, managers, and decision makers for the M&S problem domain. Have the panel assess the appropriateness of the hierarchy of indicators created. The hierarchy should be judged to determine if:

- it is sufficiently comprehensive covering all essential elements of the problem / application domain,
- it has sufficient depth providing the needed level of detail in measurement, and
- the leaf indicators are directly measurable or assessable.

Have the hierarchy revised and finalized based on the review panel recommendations.

### 2.4 Step 4: Nominal Score Sets

A leaf indicator can be assigned a crisp, fuzzy or nominal score, defined as follows.

- A *crisp score* is a single real value between 0 and 100 (e.g., 83.6).
- A *fuzzy score* is an interval of real values within the range of 0 and 100 (e.g., [67.4, 73.8]). The end points of the interval are always inclusive.
- A *nominal score* is a named score with a predefined crisp or fuzzy value.

In EE, all scores are represented as fuzzy scores (also called interval scores, numerical scores, or score ranges) on a scale between 0 and 100. A score in EE is defined in three parts:

- *Low Score*: low value of the interval score.

- *Average Score*: mid value of the interval score.
- *High Score*: high value of the interval score.

A crisp score (e.g., 85) is represented as an interval score (e.g., [85, 85]).

Table 1 shows an example set of nominal scores with predefined numerical fuzzy values or score ranges. Note that the high and low scores in adjacent score ranges are the same since the score values are real. The plus sign means “more” and implies an increase in the goodness or badness of the assessment. The minus sign means “less” and implies a decrease. The plus sign goes upward for positive assessments, and downward for negative assessments; for example, “Poor +” means more poor and “Poor -” means less poor. Therefore, “Poor +” has a lower score range than “Poor -”.

Table 1: Example Nominal Score Set

<i>Nominal Score</i>	<i>Numerical Score</i>
Excellent +	[97 .. 100]
Excellent	[94 .. 97]
Excellent -	[90 .. 94]
Good +	[87 .. 90]
Good	[84 .. 87]
Good -	[80 .. 84]
Satisfactory +	[77 .. 80]
Satisfactory	[74 .. 77]
Satisfactory -	[70 .. 74]
Poor -	[67 .. 70]
Poor	[64 .. 67]
Poor +	[60 .. 64]
Unacceptable -	[40 .. 60]
Unacceptable	[20 .. 40]
Unacceptable +	[0 .. 20]

Define nominal score set(s) at the project level. Set one nominal score set as the project default. Select a nominal score set for each leaf indicator requiring one other than the project default.

### 2.5 Step 5: Relative Criticality Weighting of Indicators

Child indicators may influence the score of their parent indicator in different amounts. For example, M&S Requirements Accuracy may effect M&S Requirements Quality much more than M&S Requirements Modifiability. Therefore, each child indicator carries a weight expressing its level of influence on its parent.

A *weight* is a fractional value between 0 and 1. The weights of the child indicators belonging to the same parent must sum to 1. Since a child indicator can influence more than one parent indicator, it can have more than one weight.

Since the weights are assigned relative to the sibling indicators to express criticality of influence on the parent

indicator, the weighting process is called *relative criticality weighting*.

Given a list of  $n$  child indicators, it is very difficult for a subject matter expert to come up with weights especially when  $n > 5$ . To facilitate the relative criticality weighting among  $n$  child indicators, we use the Analytic Hierarchy Process (AHP). Figure 2 illustrates the use of AHP for criticality weighting of the M&S Requirements Quality indicator's child indicators.

Have the Administrators and SMEs agree on child indicator weights throughout the entire hierarchy of indicators.

## 2.6 Step 6: Project Management Attributes

Have the Administrators create attributes for the leaf indicators for the purpose of project management. Example attributes include status, priority, criticality, and importance. The status attribute can be assigned a value such as "In

Progress" or "Completed" to indicate the status of evaluation with respect to a leaf indicator. Priority can be a number designating the order of executing the evaluation tasks.

Indicators can be listed in different ways, for example, by sorting with respect to criticality, then priority, then status, and then indicator name. Such listings enable the project Administrators to view the project tasks (leaf indicators) under different perspectives and to keep track of project progress.

## 2.7 Step 7: Identification of Evaluation Techniques

Scores for the leaf indicators can be determined by a variety of approaches including:

- Testing
  - Product Testing
  - Process Testing

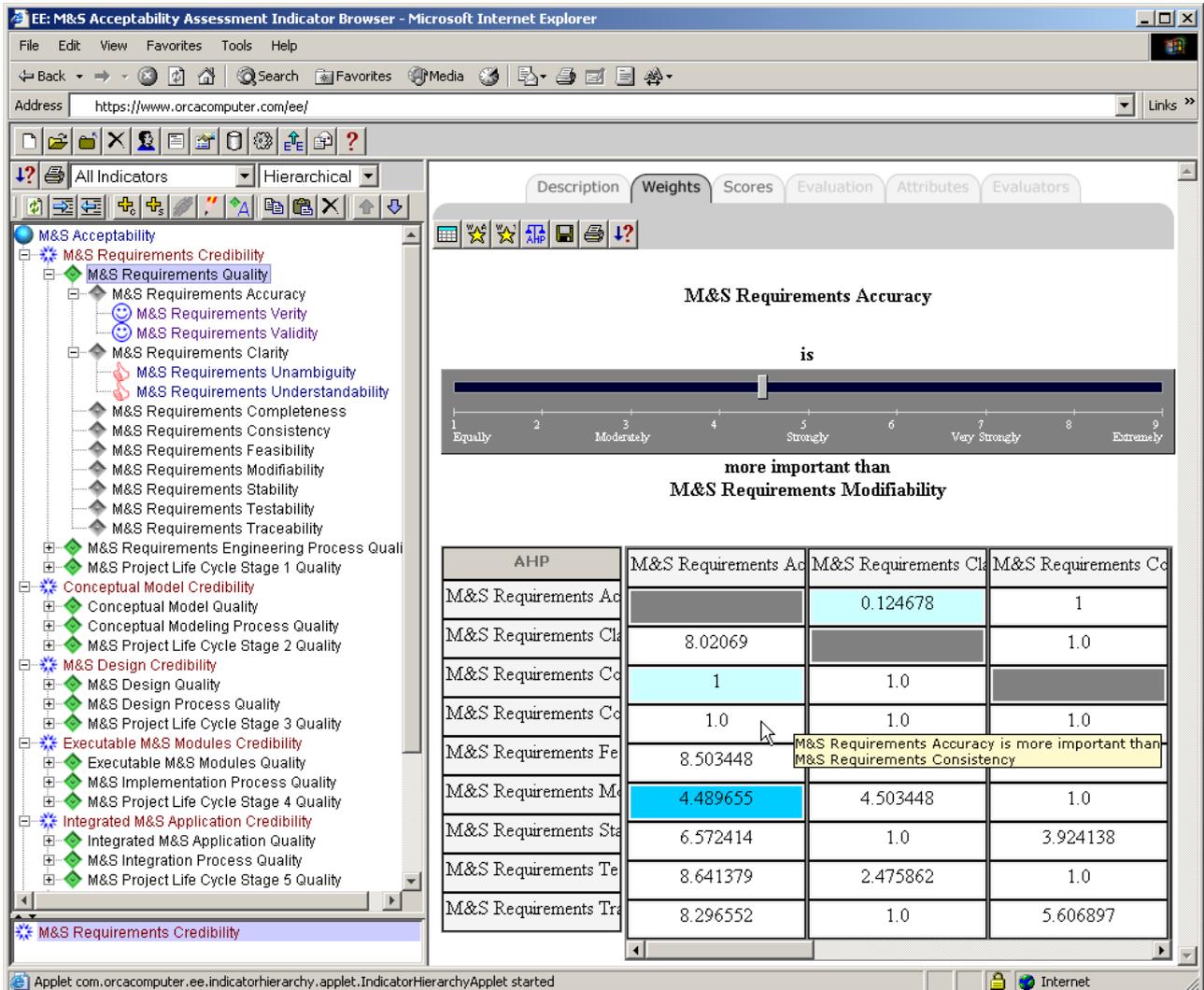


Figure 2: Relative Criticality Weighting of Indicators Using Analytic Hierarchy Process

- Field Testing
- Direct Measurement
- Analysis
  - Data Analysis
  - Graphical Analysis
  - Mathematical Analysis
  - Metrics Analysis
  - Physics Analysis
  - Probabilistic Analysis
  - Statistical Analysis
- Examination
  - By domain SMEs
  - By technical SMEs.

Under the examination approach, an SME directly determines a score for a leaf indicator based on his/her expertise, knowledge, and experience about the subject matter being evaluated. Score determination using the other approaches first requires the production of the results based on the approach used. Then, the results are interpreted and judged for goodness. The judgment is specified as a score for a leaf indicator as depicted in Figure 3.

**2.8 Step 8: Identification of Evaluators**

Identify Evaluators for the project. The Evaluators are technical people who interpret testing, direct measurement or analysis results, judge them for goodness, and specify a score for a leaf indicator. If the examination technique is used, identify technical and domain SMEs who can specify a score for a leaf indicator based on examination and using their knowledge and expertise. Have all Evaluators register as EE users.

**2.9 Step 9: Assignment of Evaluators**

Assign Evaluators to leaf indicators in the hierarchy depending on the credentials of the evaluators and/or availability of the personnel. An Evaluator may not have the

knowledge and experience to be able to score on all of the leaf indicators.

**2.10 Step 10: Relative Criticality Weighting of Evaluators**

A project Administrator may assign more than one Evaluator to assess a leaf indicator. Evaluators may have varying degrees of knowledge and experience and their opinions / evaluations may be taken into consideration with different weights.

Suppose that the *i*th Evaluator gives a score  $S_i$ , a fractional value between 0 and 100, in the evaluation of a leaf indicator. A straight average of the scores implies that each Evaluator’s evaluation is counted equally in determining a score for the leaf indicator. However, the evaluations may be counted differently based on the credentials of the evaluators. Thus, the evaluators assigned to the same leaf indicator are weighted among themselves and a weighted score is computed for the leaf indicator:

$$S = w_1 \times S_1 + w_2 \times S_2 + \dots + w_n \times S_n. \quad (1)$$

A *weight* is a fractional value between 0 and 1. The weights of the Evaluators assigned to evaluate the same leaf indicator must sum to 1.

Since the weights are assigned relative to other Evaluators to express criticality of influence on the leaf indicator score, the weighting process is called *relative criticality weighting*.

Given a list of *n* Evaluators, it is very difficult for an evaluation project Administrator to come up with weights especially when  $n > 5$ . To facilitate the relative criticality weighting among *n* Evaluators, we use AHP.

**2.11 Step 11: Evaluations**

Start the evaluation process. An evaluator interprets the results of testing, direct measurement or analysis, judge them for goodness, and specify a score for a leaf indicator. If the

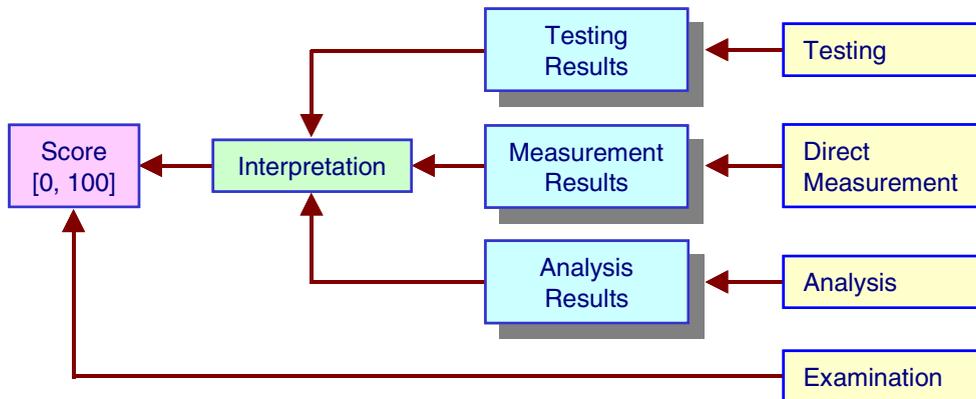


Figure 3: Leaf Indicator Score Determination

examination technique is used, the Evaluator specifies a score for a leaf indicator based on examination and using his/her knowledge and expertise.

An Evaluator is expected to document his/her rationale for a leaf indicator score using the Evaluation tab. An Evaluator can upload any kind of file including evaluation reports, multimedia files, briefing slides, and spreadsheets. The uploaded files are stored in the project file repository for access by the project users.

Once all Evaluators score on the leaf indicators, scores are aggregated by EE in a bottom-up fashion from the leaf indicators to the branch indicators and ultimately to the root indicator.

Kiviati graphs can be created in Scalable Vector Graphics (SVG) format to visualize the scores simultaneously. Figure 4 shows a Kiviati graph of the scores of seven indicators. The center of the circle represents a score of zero and the perimeter represents the perfect score 100. Each axis from zero to the perimeter 100 displays the score range of the corresponding indicator as low score, average score, and high score. The smallest area formed by the low scores represents the worst-case scenario and the largest area formed by the high scores represents the best-case scenario. The middle area formed by the average scores represents the most-likely scenario.

## 2.12 Step 12: Project Guests

Identify Guest users for the project. The Guests are people who are interested in seeing the evaluation project and its results. Have all Guests register as EE users. Set permissions to allow a particular Guest user to see only certain components of the project. You may want to hide the identity of each Evaluator in giving permissions to the Guests.

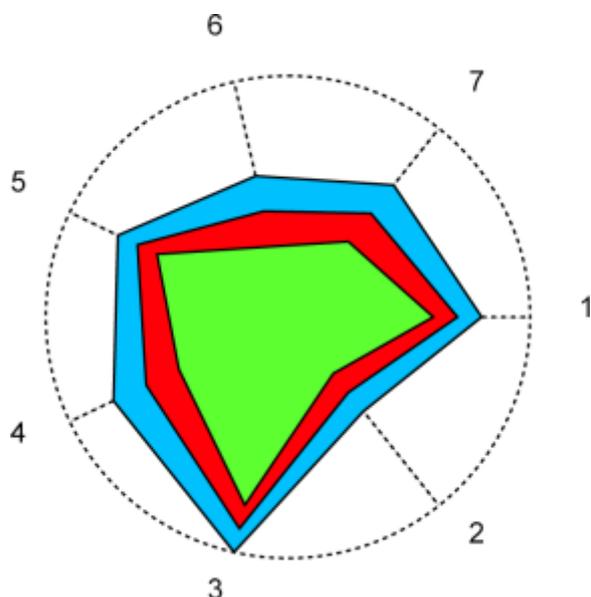


Figure 4: Kiviati Graph

## 2.13 Step 13: Project Report

Generate a project report in the Portable Document Format (PDF) based on options you specify. You can either distribute the PDF file to interested people or have them designated as project Guest users so that they can view the project and its results on the EE system online.

## 3 CONCLUDING REMARKS

Complex evaluations, such as M&S Credibility Assessment, demand rigorous collaborations among technical people, SMEs, engineers, project managers, and program managers who are geographically dispersed. EE provides the needed collaboration capabilities on the Internet through SSL with 128-bit encryption. EE's collaboration capabilities significantly reduce the VV&A cost and increase the effectiveness of VV&A.

EE uses open technology standards such as extensible markup language (XML), Extensible Stylesheet language (XSL), SVG, Dynamic Hypertext Markup Language (DHTML), and PDF. EE has its own XML markup language called EEML for project data import/export, archive/restore, and report generation. EE enables business-to-business data exchange using the XML technology.

## REFERENCES

- Balci, O. 1997. Principles of simulation model validation, verification, and testing. *Transactions of the Society for Computer Simulation International* 14 (1): 3-12.
- Balci, O. 1998. Verification, validation, and testing. In *The Handbook of Simulation*, ed. J. Banks, 335-393. New York, NY: John Wiley & Sons.
- Balci, O. 2001. A methodology for certification of modeling and simulation applications. *ACM Transactions on Modeling and Computer Simulation* 11 (4): 352-377.
- Balci, O., R. 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. Piscataway, NJ: IEEE.
- Balci, O. and S. D. Saadi. 2002. Proposed standard processes for certification of modeling and simulation applications. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. Piscataway, NJ: IEEE.
- DMSO. 2000. *DoD Verification, Validation, and Accreditation Recommended Practices Guide*, millennium ed. Alexandria, VA: Defense Modeling and Simulation Office.
- DoDI. 1996. DoD modeling and simulation verification, validation, and accreditation. Department of Defense Instruction 5000.61, Apr.

- IEEE. 1998. *IEEE Standards for Software Verification and Validation*. IEEE Standard 1012. Washington, DC: IEEE Computer Society.
- Rae, A., P. Robert, and H.-L. Hausen. 1995. *Software Evaluation for Certification: Principles, Practice, and Legal Liability*. London, UK: McGraw-Hill.
- Voas, J. M. 1999. Guest editor's introduction: certification: reducing the hidden costs of poor quality. *IEEE Software* 16 (4): 22-25.

## 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-solutions, e-business, e-commerce); M&S; and VV&A, IV&V, and certification of M&S applications and software systems. His email and Web addresses are <balci@vt.edu> and <<http://manta.cs.vt.edu/balci>>.

**ROBIN J. ADAMS** is a Software Engineer at Orca Computer, Inc. He received his B.S. degree in Computer Science from Virginia Tech in May 2001. He has been the Lead Developer of the Web-based Evaluation Environment software system at Orca Computer, Inc. since August 2000. Mr. Adams was a Lead Software Engineer for the Center for Innovation in Learning project at Virginia Tech in 1999-2000. In this project, he contributed to the development of animations to assist learning some key computer science topics. The work was later published in the *ACM Journal on Educational Resources in Computing*. His areas of expertise center on collaborative Web-based software architectures, human-computer interaction, and educational computing. He is a member of Upsilon Pi Epsilon Computer Science Honor Society, Phi Beta Kappa National Honor Society, and ACM. His email address is <Adams@OrcaComputer.com>.

**DAVID S. MYERS** is a Software Engineer at Orca Computer, Inc. He received his B.S. degree in Computer Science from Virginia Tech in May 2002. He has been contributing to the development of the Web-based Evaluation

Environment software system at Orca Computer, Inc. since September 2001. Mr. Myers has also been a Lead Software Engineer for Graphical Simulation and Modeling Environment project at Virginia Tech since 2000. Mr. Myers developed mathematical simulation models in many aspects of microelectronic design and also created several XML-based GUI tools allowing user-input and display of simulation models. The work was later published in the *Proceedings of the 14th IEEE Biennial University / Government / Industry Microelectronics (UGIM) Symposium*. His areas of expertise center on distributed and Web-based software engineering, and distributed simulation. He is currently President of Upsilon Pi Epsilon Alpha chapter of Virginia, and a member of Phi Beta Kappa National Honor Society, Golden Key National Honor Society, National Society of Collegiate Scholars, and ACM. Mr. Myers is currently working on his M.S. degree in Computer Science at Virginia Tech in the area of distributed M&S visualization. His email address is <Myers@OrcaComputer.com>.

**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>>.