# A Proposed Model for Simulation Validation Process Maturity

**S. Y. Harmon**
Zetetix
Myrtle Creek, OR  97457
*harmon@zetetix.com*

**Simone M. Youngblood**
Defense Modeling and Simulation Office
Alexandria, VA  22311
*syoungblood@dmso.mil*

This paper proposes a model of process maturity for simulation validation. The development of this model begins by recognizing validation as a process that generates information as its sole product and therefore resembles a systematic quest for truth. These characteristics distinguish the simulation validation process from other processes such as those for manufacturing or software engineering. This development then substitutes process objectivity and the properties of quality information for the difficult-to-measure qualities of truth. The properties of information quality are defined by the completeness and correctness of the information and the confidence that the information producer has that the information is complete and correct enough to serve a particular purpose. These parameters come by amalgamating the notions for information quality held by the process improvement and traditional scientific communities. The validation process takes validation criteria, the referent, the simulation's conceptual model, verification results of the intermediate development products, and the simulation results as input and produces evidence assessing the simulation's validity as output. The quality of the input information ultimately limits the quality of the evidence that the validation process produces. The proposed model of validation process maturity thus structures the validation process into six levels defined by the quality of the input information, the quality of its information products, and the objectivity of those process components that contribute to those products. At the first level, the users demand absolutely no validation of the models they apply. The second level consists entirely of face validation where the validation evidence depends entirely upon subjective sources of requirements, referent, and validity judgments. The next level improves the objectivity and quality of the representational requirements. The next two levels progressively improve the objectivity of the conceptual modeling and results validation component processes and the quality of the evidence they produce. In the final maturity level, the validation process automatically transforms informal user need statements into formal validation criteria and then applies mathematical techniques to prove conceptual model and simulation results validity. In this model, the objectivity and quality of the validation evidence increases as the level of simulation validation process maturity increases.

**Keywords:** Simulation validation, process maturity, simulation verification

## 1. Introduction

The software and systems communities have both advanced the state of their practices by applying models of capabilities to gauge the maturity of their practitioners and their processes [1, 2]. These models build capability maturity in layers that successively lead to greater end product quality. The success of these capability maturity models suggests that such a concept could promote the advancement of simulation validation practice as well.

Others have considered this problem both for simulation in general [3] and simulation verification, validation, and accreditation (VV&A) [4]. Scholten and Udink ten Cate have developed a simulation maturity model (SMM) in which they show the role of validation in promoting simulation quality [3]. Conwell, Enright, and Stutzman demonstrate how to use the Carnegie Mellon University Software Engineering Institute's (SEI's) software and system acquisition capability maturity models (SW-CMM and SA-CMM) to improve simulation development and acquisition [4]. Both of these approaches demonstrate the importance of validation to the simulation development and use

*JDMS,* Volume 2, Issue 4, October 2005   Pages 179–190

processes but do not explicitly structure validation processes themselves. On the other hand, Pace squarely addresses this structuring by arranging validation into three levels of progressively increasing completeness [5]. These levels loosely relate possible uses to the amount of validation effort. However, these levels do not specifically structure the validation processes associated with each level and do not explicitly consider process maturity. Pace does address the elements of improved validation, and indirectly validation process maturity, in later work but does not structure these processes into consecutive levels [6]. Logan and Nitta propose a series of successive validation levels that also correspond to the completeness of validation [7]. This structuring equates the quality of the validation process and its products to this notion of completeness. Oberkampf et al. extensively survey the development of validation metrics for physical and engineering models and simulations [8]. They propose necessary characteristics of validation metrics primarily focused upon assessing validity and describe desirable aspects of validation processes that take advantage of these metrics. This work implies several elements of validation process maturity but considers specifically the comparison of simulation results to the referent information, only a part of the entire validation process.

A structuring of simulation validation processes into levels of maturity can serve several purposes. These include

- Organizing validation processes so that increased effort leads to increased confidence in the validation results (implying reduced risk in using the simulation for a purpose);
- More clearly showing the return of increased investment in validation;
- Illustrating a systematic approach for moving from completely subjective validation assessments (i.e., face validation) to completely objective validation assessments by systematically attacking the high payoff areas of validation technology; and
- Identifying the information needed to support different approaches to validation.

With these goals in mind, a model of simulation validation process maturity was constructed. In constructing this model, we did not attempt to map the SEI CMM directly onto validation processes as proposed by some authors [3, 4]. Instead, we identified the basic reasoning underlying maturity matrices, in general, and the CMM, specifically, and developed the notion of validation process maturity from its first principles following an analogous reasoning path. Like the CMM, this derivation depends upon the assertion that the primary measure for improving any information process, such as simulation validation, is improvement in the quality of the information that process produces.

## 2. Validation Process Maturity Development

Like most creative processes, the elements of the validation process maturity model originated and evolved in seemingly random lurches induced by thought, discussions with colleagues, and numerous presentations to the technical community. This section organizes and summarizes the results of that three-year development activity into discussions of the model's evolution and its relationship to validation information quality, the proposed validation process maturity model itself, and the implications of that model on validation processes and products.

### *2.1 Model Evolution and Validation Quality*

The proposed model for simulation validation process maturity grew from the ideal of a parallel to the SEI SW-CMM or the System Engineering Capability Maturity Model (SE-CMM). The SW-CMM was founded upon the basic concepts of statistical process control and quality management first articulated by Shewhart [9, 10], Deming [11], and Juran [12, 13]. Radice [14, 15] applied these concepts, together with the notion of quality grids or maturity matrices from Crosby [16], to the problem of software quality. Humphrey later evolved this idea into the software maturity model [17-19] that was then matured by the SEI [1]. Many maturity models have since been proposed that build, in one way or another, upon the SEI CMMs [20, 21]. Some of these models simply adapt the CMM structure to other purposes while others return to the first principles of maturity matrices and build from there. The validation process maturity model falls into the class of the latter, a return to the first principles.

This historical perspective is important to the development of the validation process maturity model because it highlights the relationship between the SEI CMMs and total quality management (TQM). Central to TQM is the use of closed-loop control to optimize product quality (i.e., the Shewhart cycles, plan-do-check-act [9]). Humphrey organized the levels of software process maturity to parallel the systematic implementation of more sophisticated process quality control [19]. This suggests that in developing the CMM, the SEI implicitly modeled the software development process as a manufacturing process. They chose to improve that process by

- Characterizing the software development process,

- Measuring the quality of the products of that process, and
- Controlling the development process to optimize product quality.

These steps led to the natural structuring that has become one of the strengths of the CMM. However, the processes associated with validating simulations differ from those associated with manufacturing products in a few important ways. Simulation validation does not produce a shrink-wrap product. Changes in purpose or the simulation characteristics necessarily make each validation effort unique. Using the same simulation for the same purpose can justifiably and safely depend upon the original validation information. But, if the purpose or the simulation changes then a new validation product must be generated that accounts for these changes. The quality of validation products cannot be easily measured by the error reports produced by using them. This fact weakens the control loop philosophy underlying the SEI CMMs. The validation process produces information upon which decisions, and the risk assumed when making those decisions, depend. So, those processes look more like the systematic quest for true and complete information (i.e., truth) rather than the assembly line manufacturing of the same or similar product lines. Thus, the essential measure of validation information quality is its truthfulness. As validation information approaches absolute truth, its quality and utility approach maximum value to the decision maker.

These observations of the dissimilarities between software engineering and simulation validation support the conclusion that a process maturity model for validation could differ substantially from that for software or systems development. Thus, the proposed model resembles the SEI CMMs, and all of the other maturity models that have been proposed, only to the point of having layered process maturity keyed to improved product quality. Since validating simulations more resembles a search for truthful knowledge than a manufacturing process, the maturity of a validation process increases, not with increasing control of the manufacturing process as do the SEI CMMs, but rather with the increasing objectivity of the validation assessment. This discovery of an underlying notion, like that for the SEI CMMs, drove the structuring of the proposed process maturity model for simulation validation.

Although truth appears a useful measure of validation product quality, meaningfully measuring the truthfulness of information ranges from difficult to impossible. Measurements of truthfulness suffer from the same problems as measurements of randomness. In the limit, complete truthfulness and complete randomness cannot be formally proven, only invalidated. In order to overcome this problem, we have assumed that the truth of an information product depends upon the objectivity of the process producing it and the quality of that information. Here, we adopted a scientific model of truth and, from that, asserted that a purely objective process (i.e., one that produces results independent of its observers) is more likely to supply truthful information than one incorporating subjectivity in the assessment process. While considerable historical empirical evidence supports the veracity of this assumption, no formal argument exists to justify its soundness. Knowledge of the characteristics of information quality can strengthen this development.

Before proceeding to consider validation information quality, we must define what we mean by simulation validity:

> *Simulation validity* – the simulation's representation of its simuland (i.e., the thing that a simulation is simulating) is fit enough to serve a particular purpose.

By fit enough, we mean that the simulation represents the simuland with sufficient fidelity to serve a user's purpose. We interpret fidelity as defined in the SISO fidelity conceptual framework [22]. This definition implies that any assessment of simulation validity must consider all of the aspects of representational quality and not just accuracy. This definition says nothing about the capabilities normally addressed by conventional systems engineering (e.g., human factors, reliability, support requirements). This is not to say that these factors are unimportant to a practical simulation but only that the simulation's representational capabilities distinguish it from all other system types. With this understanding of the meaning of validity, we can safely proceed to examining the properties of information quality in the context of simulation validity.

People have given considerable thought to the issue of information quality, especially in the context of evaluating the information presented on websites [23]. Many of these sources seem to confuse the issues of information quality with those of the credibility of the source. While credibility and truth can correlate, they are not identical. Nevertheless, the consensus of opinion on the attributes of information quality seems to include the following aspects [24-27]:

- Objectivity,
- Repeatability,
- Timeliness,
- Completeness, and
- Accuracy.

We have already argued for the importance of objectivity and repeatability in simulation validity assessments, but these characteristics do not lend themselves to ready measurement and result more from the process used to obtain the information than from the information itself [28]. Timeliness is important in simulation validation in three regards:

- Relevance to a use,
- Relevance to the simuland and its referent, and
- Relevance in the face of the simulation's change.

Validity assessments inherently incorporate the relevance to a use by determining a simulation's fitness for a particular purpose. Validity assessments can become dated in the face of change when the capabilities of the simulation change or when its simuland (and its referent) changes. If the simulation remains static then an old assessment should contain identical information to a more recent assessment for the same purpose. Changes to an evolving simuland or to the knowledge about it could make a simulation's representation inaccurate. Thus, some measure of a validity assessment's timeliness can be valuable so that someone considering using that information could determine its currency.

Physical experimentation typically attaches the notions of uncertainty to measurements of any physical property state [29]. The error bars commonly used when plotting physical measurements or data derived from measurements explicitly reflect the uncertainty associated with the measurements and implicitly suggest the confidence associated with these error intervals—actually, the confidence interval delimits the probability that the actual mean lies within the error bound [30]. Oberkampf et al. explicitly link the determination of uncertainty to the validation assessments of physical and engineering simulations [8]. Logan and Nitta also propose evaluating the quality of simulation validation assessments through the information that they provide on simulation error and confidence. They argue for the need for confidence in relating simulation validity to the risk of using the simulation for a purpose [31].

We therefore suggest that a complete characterization of simulation representational capabilities will contain the following information:

- Enumeration of the entities, properties, and dependencies that the simulation represents;
- Characterization of the upper bounds of the errors at which the simulation represents the states of the entities that it represents; and
- Estimates of the confidences with which those errors are known.

This proposition implies that the process of validating a simulation begins by determining the minimum representational capabilities sufficient to serve a purpose in these terms. The process then measures the actual representational capabilities of the subject simulation in these terms. Finally, a validation agent can assess the simulation's validity for the purpose by comparing the simulation's observed representational capabilities against the representational requirements expressed in the validation criteria. In the simplest case, a simulation possesses sufficient capabilities to completely meet the purpose's validation criteria. More likely, differences will exist between the requirements and the actual capabilities. Then the validation agent can only identify those differences and let either the user or their representative (e.g., the accreditation agent) decide how those differences affect the simulation's applicability to a purpose.

The idea of expecting that simulation requirements will be stated in such detailed terms may appear unrealistic considering the present state of requirements engineering for simulations. However, these requirements need only define the characteristics of the input available and the output needed to address the user's purpose. A simulation may involve many more entity types, properties, and dependencies in order to represent sufficient causality to achieve the required accuracy of the output from a given set of input. However, the actual representational requirements need only specify that accuracy and not the accuracies of all of the components necessary to achieve it explicitly. This greatly simplifies requirements specification as well as simulation validation.

Therefore, any simulation validation assessment will identify the following:

- Entities, properties, and dependencies needed to produce the output that the user requires to serve their purpose;
- Maximum errors in the output attribute values tolerable to the user and the property value ranges over which they need this maximum tolerable error to apply;
- Confidences that the user requires to reduce the use risk to achieve their goals;
- Entities, properties, and dependencies that the simulation actually represents;
- Maximum errors in the property values that the simulation produces as output and the ranges of over which those maximum errors can be guaranteed for given input conditions;
- Confidences in the observations of the output characteristics; and
- Differences between the requirements and the simulation capabilities that may exist.

The validation agent must get, as input, or interpret, from the input they get, the representational requirements and must observe or measure the simulation's actual representational capabilities. In a sense, a simulation creates an artificial causality through its execution of its dependencies. Assessing the validity of a simulation determines whether its representation of causality adequately serves a purpose. Ideally, an independent observer should be able to obtain this same information through measurements on the same simulation given the same purpose.

Through this line of consideration, we can define the quality of simulation validation information by its possession of the following three necessary components.

- *Completeness* – A validation assessment must determine if the simulation represents all of the things that the user requires to pursue their purpose. It should also identify the required things that the simulation does not represent. Identifying those things that the simulation does represent but that the user does not specify as needed is optional. In this case, the user or their representative should decide whether that information about the simulation's capabilities is important to their acceptance decisions.
- *Correctness* – A validation assessment must identify if the simulation's representation matches the simuland's behavior sufficiently to adequately serve the user's purposes. It should also identify those attributes where the simulation's representation has excessive error. Again, identifying those areas where the simulation has more accuracy than needed is optional. Unnecessary representational capability can increase the costs of use, so the user or their representative should determine if they need this additional information. Further, computation of representational errors requires a referent to define the standard of the simuland's behavior. Referent choice can significantly affect the credibility of the validation assessment. Therefore, the validation assessment must identify the referent used to support error computations.
- *Confidence* – A validation assessment must explicitly characterize the confidence that the user can place in its information, particularly the error estimates. It should assign confidence intervals to each error estimate in such a way as to represent all of the sources of uncertainty associated with the validation measurements (i.e., not with the representational requirements as they come from the user). It should also identify areas where either the simulation or the validation assessment cannot provide sufficient confidence to meet the user's requirements and suggest the means through which to increase that confidence (e.g., performing more results testing, improving the referent information).

In thinking about accuracy and error, one is really saying that if I give the simulation input (in the form of various data sets; e.g., terrain databases, entity characteristics, user input) with a certain maximum error then I can expect that the error of the output (as compared with the referent) will not exceed specified maxima. The absolute error (again, as compared with the referent) can only be determined for a completely self-contained simulation (i.e., one that does not

**Table 1.** Proposed tiers of simulation validation

| Tier of Validation | Supporting Information | Informal Validity Statement |
|---|---|---|
| 0 | Nothing | I have no idea. |
| 1 | Simple statement of validity | It works; trust me. |
| 2 | Required entities and attributes compared against the entities and attributes that the simulation represents | It represents the right entities and attributes. |
| 3 | Required entities, attributes, and dependencies compared against entities, attributes, and dependencies represented | It does the right things; its representations are complete enough. |
| 4 | Required entities, attributes, dependencies, and dependency errors compared against entities, attributes, and dependencies represented and representation errors | For what it does, its representations are accurate enough. |
| 5 | Required entities, attributes, dependencies, dependency errors, and confidences in assessment compared against represented entities, attributes, dependencies, representation errors, and assessment confidences | I'm this confident that this simulation is valid. |

require any externally derived data or user input). In that case, the contributions of all of the resident data contributing to the simulation's error are well known.

Confidence has many interpretations [30, 31]. We intend that the assessed confidences reflect the totality of uncertainties associated with making a validation assessment. For instance, measures of confidence are particularly necessary when the simulation uses stochastic approximations of causality. These representations substitute randomness to economically represent complex causal situations. In stochastic representations, the accuracy of the representation can only be known to a certain prescribed confidence. This non-determinism built into the simulation complicates the measurement of its errors. Further, confidence can measure the degree of confidence that one has in the truthfulness of the validation assessment. Several factors can decrease this confidence from unity including the incompleteness of coverage of the behavior space, incompleteness of the requirements, and incompleteness of the referent. Logan and Nitta and Oberkampf et al. describe more about the influences of these different sources of stochasticity [8, 31]. The coverage of the simulation behavior space becomes important when the validation assessment can only sample some fraction of that space, which is most often the case. As a result, the assessed confidence should also include the effects of the uncertainty implied by incomplete results sampling of a simulation's behavior space as well as any uncertainty associated with the referent and the processes for comparing the referent with the simulation results to assess simulation accuracy [8].

In addition, a validation assessment may need to identify the sources of information upon which it depends for credibility purposes.

Recognizing these attributes of a validation assessment enables our proposing the tiers of validation shown in Table 1. In this structuring, we argue that successively adding information to the validation assessment, as described above, requires more effort but also improves the value of the assessment to the user. For example, assessment of a simulation's functionality and error limits identifies what the user can do with the simulation while still operating beneath specified error limits. Identifying the confidences contributes to the user's assessments of impact risk. Therefore, the increasing tiers of validation conform to increasing effort as well as increasing value to the user.

No validation of the simulation is done at Tier 0. As in the validation process maturity model, this tier establishes the baseline for validation tiers [1].

At Tier 1, one or more persons with credibility to the user have evaluated the simulation's behavior informally and determined that it suits the user's purposes. They reflect this assessment with a simple statement that the simulation is either valid or invalid. The strength of this declaration of validity depends entirely upon the credibility of the individuals offering the assessment to the simulation users.

At Tier 2, someone has examined the simulation to determine that it represents enough of the entities and their properties to accomplish the user's purpose. This suggests that someone has articulated the user's representational requirements in terms of the needed entities and their properties. This tier illuminates a static view of the simulation's capabilities and its validity at this state of assessment.

At Tier 3, someone has examined the simulation to determine that the attributes of the entities change in a way sufficient to meet the user's needs; i.e., that the simulation represents enough of the dependencies between the entity properties to satisfy the user's needs. This assumes that someone has interpreted the user's representational needs enough to identify the dependencies required to meet those needs. This begins the process of describing the simulation's dynamic characteristics and their relevance to its validity.

At Tier 4, someone has examined the simulation to determine that the accuracies of the representation and the ranges of state over which those accuracies apply are sufficient to meet the user's needs. In this case, the accuracies refer to those of the property representations but ultimately of the dependencies between properties. This assumes that someone has interpreted the user's needs to determine the minimum accuracies that they require (or, more properly, the maximum errors that they can tolerate) to achieve their goals. This also assumes that someone has identified a credible referent against which to measure simulation errors and selected the processes to meaningfully compare referent against simulation capabilities as described by Oberkampf et al. [8].

At Tier 5, someone has evaluated the sources of uncertainty and computed the confidence with which the simulation's behavior is understood and characterized by the functional and error assessments. Several sources contribute uncertainty to a validity assessment including the requirements, the simulation evaluation, and the referent. Further, techniques need to be applied when comparing the observed simulation capabilities against the referent that produce estimates of confidence in the results of that comparison [8].

### 2.2 Proposed Process Maturity Model

The proposed model, shown in Table 2, organizes validation activities into six levels of process maturity into which the five tiers of validation product quality map. Each succeeding level enables increasingly

**Table 2.** Levels of simulation validation process maturity

| Level | Validation Criteria | Referent | Conceptual Model (CM) | Development Products | Simulation Results |
|---|---|---|---|---|---|
| 0 | None derived | None chosen | None constructed | Verified enough to support development | Not validated at all |
| 1 | Represented by subject-matter expert (SME) opinion | Represented by SME opinion | None constructed | Verified enough to support development | Validated by SME observing simulation results |
| 2 | Defined from user statements in terms of entities represented, their attributes, and the dependencies between those dependencies | Represented solely by SME opinion | Validated against the validation criteria by the SME | Verified against the CM inventory | Validated by SME against the validation criteria |
| 3 | Defined from user statements in terms including attribute ranges, domains, and errors | Derived from a single source | Validated by objective party from validation criteria and referent | Verified against the CM | Evaluated by objective party from validation criteria and referent |
| 4 | Defined from user statements in terms of entities, attributes, dependencies, ranges, domains, errors, and confidences | Sampled from multiple independent sources and correlated statistically with estimates of uncertainties | Validated by objective party from validation criteria and referent; analyzed to suggest results sampling space and to estimate the confidence associated with that sampling | Verified against the CM; provides information to guide results sampling and to estimate the confidence associated with that sampling | Sampled from guidance developed from CM and verification results analysis; validated by objective party from validation criteria and referent |
| 5 | Formally derived from user statements using causality arguments | Rigorously derived from multiple independent sources and characterized statistically with estimates of uncertainties | Formally stated and validated automatically from validation criteria and referent; analyzed to define results validation sample space | Verified against CM and used to define results validation sample space and the confidence associated with that sampling | Automatically sampled from guidance defined from CM and verification results analysis; validated automatically from validation criteria and referent |

objective validation assessments. Much of the technology required to achieve the higher levels is unproven, underdeveloped, or undeveloped. The maturity levels defined by Table 2 do not define levels of simulation credibility but may contribute to that concept if the simulation users subscribe to its rationale.

Table 2 defines validation activities in terms of the validation-related operations performed upon the information artifacts associated with simulation development and use. These include

1) Validation criteria,
2) Referents,
3) Conceptual models,
4) Development products, and
5) Simulation results.

The development products include such intermediate implementation artifacts as detailed design information, software development products, and implementation components.

Level 0 of validation process maturity assumes that no validation is performed and is included for completeness. A validation process at this maturity level produces the weakest form of simulation validation evidence: none. It generates assessments of validity corresponding to Tier 0 in Table 1.

Level 1 depends entirely upon face validation, because no other reliable information about requirements or the referent is available. The credibility of this level of process maturity, as well as the validation assessments it generates, depends completely upon the faces doing the validating. Many past and current efforts to validate simulations of complex phenomena (e.g., theater-level warfare) rely upon processes at this level of maturity. This level generates assessments of validity at Tier 1 in Table 1.

Level 2 begins the process of removing the dependence upon the SME and, thus, reducing the subjectivity associated with the complete reliance upon SME assessments at a Level 1 maturity. While it still depends upon the SME to interpret requirements from user needs statements, the resulting validation criteria describing these requirements are stated in the observable terms of the

- Objects that the simulation should represent,
- Properties characterizing and distinguishing those objects, and
- Dependencies that couple the behavior (i.e., state changes over time) of those properties.

These representational characteristics are the first elements of simulation fidelity [22, 32]. These attributes define a simulation's representational completeness. This level takes the first step in achieving more objective validation by getting requirements stated in terms of observable validation criteria. However, the absence of validation criteria that define required representational accuracies and a referent against which to measure those accuracies prevents the user from removing the SME entirely from evaluating the validity of both the conceptual model and the simulation results. The determination of validation criteria from user needs assumes that the user verifies the adequacy and correctness of the resulting criteria in this level and in every subsequent level. This step contributes to building the user's belief in validation assessment credibility as well as ensuring the correspondence between validation criteria and actual user needs. A Level 2 process produces information at the Tier 3 of validity in Table 1, skipping over Tier 2 by producing validity assessments including both static and dynamic information. In effect, the Level 2 process maturity could be partitioned into Level 2a and Level 2b with one including only the assessment of a simulation's representation of state and the following partition evaluating both the representation of state and behavior.

Level 3 continues with a richer description of the validation criteria, which include

- Domains,
- Ranges, and
- Tolerable errors

for each dependency represented. These attributes begin to define the quantitative elements of simulation fidelity [22, 32]. The quantitative measurement of simulation representational errors requires using an objective referent against which to measure simulation representational errors. While, at this level of maturity, the referent can come from a single information source, that source must provide knowledge that is independent of the user's needs and of the particulars of the simulation being examined. This step partitions the subjectivity of SME sources, if chosen as referents, from the processes of interpreting user needs and assessing simulation validity. This may not directly improve process objectivity by itself but does make gauging the impacts of subjectivity upon the process easier. This referent must define the ranges and domains over which its measures of accuracy pertain. This, in effect, characterizes the conditions under which the referent information is useful. The availability of validation criteria describing tolerable errors and an independent referent make validating the conceptual model and simulation results with an objective party (i.e., not an SME) possible. Level 3 permits referents derived from a single source of information. This information is sufficient if uncertainties are not estimated. When one assesses the confidences then multiple correlated and statistically characterized referents are needed.

Level 4 adds additional description to the validation criteria as the confidences that the user desires in the information produced by the validation assessment.

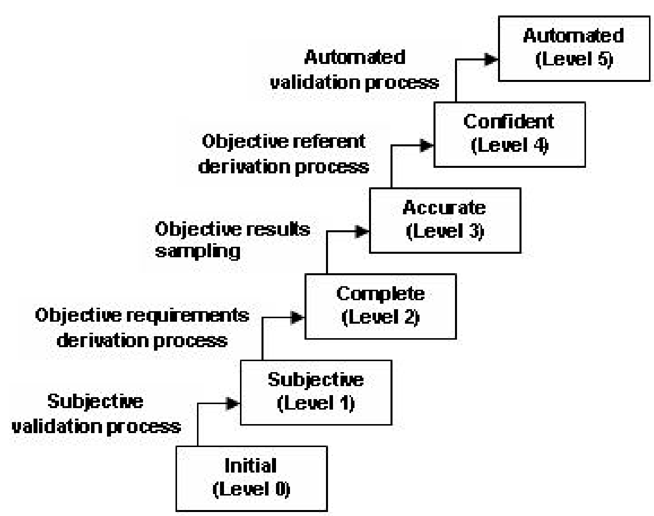


**Figure 1.** Levels of validation process maturity and the processes needed to improve maturity

**Table 3.** Effects of validation process maturity on judges of validity, validation criteria, and confidence estimates

| Level | Judge of Validity | Validation Criteria | Confidence Estimates |
|---|---|---|---|
| 0 | None | None | No |
| 1 | SME | SME knowledge of user needs | No |
| 2 | SME | Explicit specification of needed entities, entity properties, and behaviors together with SME knowledge of user needs for accuracy | No |
| 3 | Independent observer | Explicit specification of needed entities, entity properties, behaviors, minimum needed accuracies, and property ranges over which minimum accuracies can be guaranteed | No |
| 4 | Independent observer | Explicit specification of needed entities, entity properties, behaviors, minimum needed accuracies, property ranges over which minimum accuracies can be guaranteed, and desired confidences in validation evidence | Yes |
| 5 | Formal proof | Formal specification of needed entities, entity properties, behaviors, minimum needed accuracies, property ranges over which minimum accuracies can be guaranteed, and desired confidences in validation evidence | Yes |

These desired confidences can be derived from the tolerable levels of risk that a user incurs from applying the simulation to their problem set. This maturity level also requires a referent that is independent of a single SME to compensate for the human's unreliable innate ability to make accurate quantitative estimates [33]. This requires constructing the referent from multiple independent sources, correlating the information supplied by those sources, and then statistically characterizing the variances associated with the correlated result. These variances contribute to the estimations of uncertainty in the validation assessment results as described by Oberkampf et al. [8]. This maturity level further improves the conceptual model validation and, in turn, the utility of the verification activities to validation. This level also adds steps that analyze the conceptual model and the verification results to identify those parts of the simulation behavior space that should be sampled for results validation. This step improves confidence in the results validation and strengthens the ties between conceptual model validation, verification, and results validation. The referent at this level comes from multiple disparate, but correlated, sources to improve its independence from the specific sources and to reduce and better characterize its uncertainties. A Level 4 of process maturity produces validity at Tier 5 in Table 1.

In Level 5, rigorous derivation of a provably necessary and sufficient set of validation criteria enables automatic validation of a formal conceptual model. This provability comes from the mathematical arguments for abstracting the causal chains necessary to accomplish the simulation's purposes. This step, along with the results from analyzing the conceptual model and verification results, also permits automatic collection and validation of the simulation results. This emphasis upon mathematical techniques requires a strong referent that has been formally abstracted from the available knowledge. This degree of automation provides the greatest objectivity but assumes the correctness of the mathematics upon which that automation is based. Further, for the resulting validation assessments to be credible, the user must also believe in the sufficiency and correctness of those mathematical techniques and their application. While the focus of this level appears to be automation of the validation process, it is not. The ability to automate simply belies the degree to which the mathematical mechanisms exist to facilitate automation. These mathematical techniques could as easily be applied manually as through automation. Nevertheless, the formal arguments and mathematical techniques needed to support this level of validation process maturity have not been sufficiently developed today.

Repeatability really describes the attributes of a process. It implies that any assessment produced by the process is independent of the specific people executing that process. In other words, two people executing the process correctly on the same simulation for the same purpose will produce identical assessments of validity. In Table 2, we have assumed that the process' repeatability improves with the increasing level of objectivity.

Figure 1 illustrates the proposed validation process maturity model as a set of six states with transitions labeled by the essential process improvements that make moving from one level of process maturity to another possible. Table 3 summarizes the central characteristics of each process maturity level.

## 2.3 Model Implications

The SEI CMMs emphasize a progression toward product quality in their levels. The structuring of validation effort presented in this paper adopts a similar emphasis but rather on the quality of the information produced by those validation processes. Decision makers employing simulation in their decision processes depend upon the quality of the information from those simulations. Simulation information quality, in turn, depends upon the quality of the validation information. The approach underlying this model of validation process maturity relies upon two key assumptions:

1) The quality of validation information depends upon its truthfulness and completeness, and improved truthfulness and completeness can only be achieved through improved objectivity; and
2) Reliably improving validation process objectivity requires understanding the fundamentals of that process.

The emphasis upon objectivity and fundamentals distinguishes this approach from that suggested by Pace [6]. Pace stresses the importance of quantitativity and reliance upon "real-world" referents in validation. Both of these properties are undoubtedly important to validation, but many aspects of simulation representations are not naturally quantifiable, and real-world referents can be as poorly understood as the qualitative ones described by Pace [6]. These aspects will neither necessarily improve nor ensure the truthfulness and completeness of the validation information. Poorly understood attempts to quantify information or statistically characterize information can lead to a false sense of improvement where the information truth may have actually been degraded.

The proposed structuring of the validation process maturity model has implications for the conceptual model content, the decision risk, and effort cost.

As the level of validation process maturity increases, the dependence upon the conceptual model information increases. The conceptual model represents the first development product that provides sufficient information to validate against the validation criteria. Once validated, the conceptual model establishes the touchstone against which to verify all successive development products. The inner three levels of process maturity systematically build the information the conceptual model contains:

- Level 2 specifies the objects, properties, and dependencies that the simulation will represent;
- Level 3 specifies the objects, properties, and dependencies as well as the ranges, domains, and accuracies of those dependencies that the simulation will represent; and
- Level 4 specifies the objects, properties, dependencies, dependency ranges, dependency domains, and dependent variable accuracies as well as the uncertainties associated with the simulation representation.

Specification completeness of the conceptual model parallels the specifications of the validation criteria because those create the standard against which the conceptual model is validated. Increased specificity in the conceptual model also makes its information more useful for tailoring verification and results validation activities. This process maturity development assumes that decision risk is controlled by the completeness and correctness of the knowledge available to the decision maker, in particular that knowledge generated by the simulation. Thus, the likelihood of making a wrong decision when given correct and complete information is assumed to be zero. This assumption stands upon the additional assumption that the decision maker has complete ability to control the problem situation if given sufficient information. This final assumption is seldom, if ever, true but presents an ideal within which to solve the simulation validation problem. Ultimately, the decision maker must determine what risk they are willing to assume when using simulation results; the validation process contributes information to making that risk assessment. Validation assessments produced by mature validation processes simplify that risk assessment and improve its accuracy over information from immature processes.

Validation effort cost is loosely coupled to validation process maturity. Increasing the resolution of the validation assessment increases the labor needed to collect, correlate, and analyze validation-related information. Therefore, as the maturity of the validation process increases, the labor to exercise it and the costs associated with that labor also increase monotonically. However, this correlation breaks at Level 5 maturity. At this maturity level, the process shifts much of the validation process labor to automation. Human effort may still be required to collect information, but machines perform all of the correlation and analysis. This step may actually decrease the cost of increasing validation process maturity when sufficient tools are available. This cost argument for improved validation process maturity neglects the costs associated with building the validation referent.

Assuming no costs associated with building the referent constitutes a very large assumption and potential Achilles heel of this process model. Assembling an adequate referent can easily exceed the costs of executing the validation process, even at the highest levels of effort. Further, assembling an adequate referent may be very difficult or impossible in some cases. The lack of an adequate referent will contribute uncertainty to the validation process and can limit the value gained from exercising the higher process levels. However, a single referent can serve many validation efforts examining the representations of the same or suitably similar simulands. This development has not examined the issues associated with simulation referents sufficiently, and more work will be done in this important area in the future.

## 3. Conclusions

This paper proposes a model of validation process maturity structured into six levels that successively increase validation process objectivity and improve validation assessment quality. However, structuring the validation problem does not make simulation validation any easier. It does take a step toward clarifying the dependence of decision risk upon rationally managed validation effort. As a result, like the SEI CMMs, this structuring more clearly depicts where an application requires more scrupulous validation attention. This organization of validation process maturity is also largely independent of simulation and use specifics as well as any particular techniques that one might apply to the validation problem. As an unexpected consequence, this arrangement illuminates the future of validation. It suggests that developing a clearer understanding of validation fundamentals can both improve validation results, with a commensurate reduction in decision risk, and reduce validation effort cost.

The validation process maturity model suffers from the same ills as do most process maturity models and is therefore subject to the same criticisms. While great care has been exercised to choose the levels and allocate their functions, one could still argue that the structuring particulars appear arbitrary and lack formality. Fraser et al. argue that rigorous development of maturity matrices is very difficult and that some compromise between rigor and utility must be struck to produce a useful and usable tool [20]. Further, the specific number of levels seems arbitrary. Again, Fraser et al. point out that this is true of all maturity matrices, CMMs included [20]. The purpose of this effort was to develop a useful tool founded upon sound reasoning to aid validation practitioners. In some cases, strict formality was abandoned in favor of completing a useful product. Only further application will tell whether the validation process maturity model achieves this goal.

Three efforts are currently underway that apply the validation process maturity model. These efforts differ in their user sets, use domains, use risks, and simulation characteristics. One proof of concept is supporting the validation of simulations for experimentation. The other two support validation of simulations for analysis and training applications. At this point in these efforts, the validation process maturity model has needed very little revision to provide useful insight into the participating validation processes. However, many lessons will be learned from these efforts that will be reported in later publications on this topic.

## 4. References

[1] Paulk, M. C., C. V. Weber, W. Curtis, and M.B. Chrissis, eds. *The Capability Maturity Model: Guidelines for Improving the Software*. Reading, MA: Addison-Wesley Publishing Company, 1995.

[2] Bate, R., et al. *A Systems Engineering Capability Maturity Model*. Version 1.1. CMU/SEI-95-MM-03. Pittsburgh: Carnegie Mellon University Software Engineering Institute, November 1995.

[3] Scholten, H., and A. J. Udink ten Cate. “Quality Assessment of the Simulation Modeling Process.” *Computers and Electronics in Agriculture* 22 (1999): 199–208.

[4] Conwell, C. L., R. Enright, and M. A. Stutzman. “Capability Maturity Models Support of Modeling and Simulation Verification, Validation, and Accreditation.” In *Proc. 32nd SCS Winter Simulation Conference*, 819–828. 2000.

[5] Pace, D. K. “An Aspect of Simulation Cost.” *Phalanx* 30, no. 1 (March 1997): 12–15.

[6] Pace, D. K. “Validation Elaboration.” In *Proc. 2002 SCS Summer Simulation Conference*. np. 2002.

[7] Logan, R. W., and C. K. Nitta. *Verification and Validation (V&V) Guidelines and Quantitative Reliability Confidence (QRC): Basis for an Investment Strategy.* UCRL-2002x0266. Livermore, CA: University of California, Lawrence Livermore National Laboratory, August 2002.

[8] Oberkampf, W. L., T. G. Trucano, and C. Hirsch. *Verification, Validation and Predictive Capability in Computational Engineering and Physics.* SAND Report SAND 2003-3769. Albuquerque, NM: Sandia National Laboratories, February 2003.

[9] Shewhart, W. A. *Economic Control of Quality of Manufactured Product.* Milwaukee, WI: American Society of Quality (ASQ) Quality Press, January 1980.

[10] Shewhart, W. A. *Statistical Methods from the Viewpoint of Quality Control.* New York: Dover Publications, November 1990.

[11] Deming, W. E. Out of the Crisis. Cambridge, MA: MIT Center for Advanced Engineering Study, 1986.

[12] Juran, J. M. *Juran on Planning for Quality.* New York: Macmillan, 1988.

[13 Juran, J. M. *Juran on Leadership for Quality.* New York: The Free Press, 1989.

[14] Radice, R. A., N. K. Roth, A. C. O’Hara Jr., and W. A. Clarfella. “A Programming Process Architecture.” *IBM Systems Journal* 24, no. 2 (1985): 79–90.

[15] Radice, R. A., J. T. Harding, P. E. Munnis, and R. W. Phillips. "A Programming Process Study." *IBM Systems Journal* 24, no. 2 (1985): 91–101.

[16] Crosby, P. B. *Quality is Free.* New York: McGraw-Hill, 1979.

[17] Humphrey, W. S. "Characterizing the Software Process: A Maturity Framework." *IEEE Software* 5, no. 2 (March 1988): 73–79.

[18] Humphrey, W. S. *Managing the Software Process.* Reading, MA: Addison-Wesley, 1989.

[19] Humphrey, W. S. "Software and the Factory Paradigm." *Software Engineering Journal* 6, no. 5 (September 1991): 370–376.

[20] Fraser, P., J. Moultrie, and M. Gregory. "The Use of Maturity Models/Grids as a Tool in Assessing Product Development Capability." In *Proc. IEEE Int. Engineering Management Conference,* August 2002. np. 2002.

[21] U. S. Department of Defense. *C4ISR Architecture Working Group. Levels of Information System Interoperability (LISI).* Washington, DC. March, 30, 1988.

[22] Gross, D. C., et al. "Report from the Fidelity Implementation Study Group." In *Proc. Spring 1999 Simulation Interoperability Workshop.* Paper #S99-SIW-167. np. 1999.

[23] Smith, A. "Evaluation of Information Sources." In *The World-Wide Web Virtual Library.* Wellington, New Zealand: Victoria University of Wellington, October 23, 2003. (online at http://www.vuw.ac.nz/staff/alastair_smith/evaln/evaln.htm)

[24] Tyburski, G. "Assessing the Quality of Information at a Web Site." In *The Virtual Chase.* Washington, D.C.: Ballard, Spahr, Andrews and Ingersoll, LLP, February 13, 2003. (online at http://www.virtualchase.com/howto/assess_quality.html)

[25] Showme Media, Ltd., Quick. *The Quality Information Checklist.* London, U.K.: Health Development Agency, August 1999. (online at http://www.quick.org.uk/)

[26] Fenton, S. J. *Information Quality: Is The Truth Out There?* Chapel Hill: University of North Carolina, May 29, 1997. (online at http://www.ils.unc.edu/~fents/310/)

[27] Adams, S. "Information Quality, Liability, and Corrections." *E-Media Live* 27, no. 5 (September/October 2003) np. (online at http://www.infotoday.com/online/sep03/ adams.shtml)

[28] Harmon, S. Y., and S. M. Youngblood. "A Proposed Model for Simulation Validation Process Maturity." In *Proc. SISO Spring 2003 Simulation Interoperability Workshop,* np. 2003.

[29] Youden, W. J. *Experimentation and Measurement.* Mineola, NY: Dover Publications, 1998.

[30] Mandel, J. *The Statistical Analysis of Experimental Data.* New York: Dover Publications, 1964.

[31] Logan, R. W., and C. K. Nitta. *Validation, Uncertainty and Quantitative Reliability at Confidence (QRC).* AIAA Paper No. 2003-1337, January 2003.

[32] Harmon, S. Y., and D. C. Gross. "On Fidelity and Related Concepts." In *Fidelity Interim Study Group Report.* Appendix B. Orlando, FL: Simulation Interoperability Standards Organization, 1998.

[33] Tourangeau, R., L. J. Rips, and K. Rasinski. *The Psychology of Survey Response.* Cambridge, England: Cambridge University Press, 2000.

Note: The URLs provided in these references are current as of December 2005.

## Acknowledgements

The Validation, Verification, and Accreditation Program of the Defense Modeling and Simulation Office generously supported this work. The authors profoundly thank Dr. Dirk Brade, Dr. David Gross, Dr. Roger Logan, Mr. Michael Metz, Dr. Dale Pace, Ms. Jennifer Park, Dr. Timothy Trucano, Mr. Robert Senko, Ms. Marcy Stutzman, and the reviewers for their insightful and helpful comments and suggestions. The authors are especially grateful to Dr. Pace for supplying pointers to his and others' work on this problem.

## Author Biographies

***Scott Harmon*** *is owner of Zetetix, a small business specializing in the modeling of complex information systems. Mr. Harmon has been developing rigorous techniques for the validation of simulation federations and human behavior representations. He has a B.S. in Physics and an M.S. in Mechanical Engineering, both from Arizona State University.*

***Simone Youngblood*** *serves as the DMSO Modeling and Simulations Verification, Validation, and Accreditation Technical Director and Education Program Manager. Her primary responsibilities include defining and implementing the DoD VV&A and education programs which include the development of DoD VV&A policy and guidance; serving as an authoritative source on VV&A issues, providing instruction/guidance to numerous individuals and DoD programs; and participating in numerous VV&A activities within the DoD and NATO communities, including the DMSO VV&A Technical Working Group, the Simulation Interoperability Workshop, and the Society for Modeling and Simulation. She has a B.A. in Mathematics and a B.S. in Computer Science, both from Fitchburg State College, and an M.S. in Computer Science from The Johns Hopkins University.*