TOWARDS A UNIFIED THEORY OF VALIDATION

Lisa Jean Bair   Andreas Tolk

Lisa Jean Bair Analytics   SimIS
4211 Monarch Way, Suite 110   200 High Street, Suite 305
Norfolk, VA 23508 USA   Portsmouth, VA 23704 USA

ABSTRACT

The modeling and simulation (M&S) literature is rich with procedures, standards, conceptual frameworks, and M&S-related theories that provide frames of reference for M&S validation. However, these works are disjoint, leading to inconsistencies: common terms used differently, differing terms for similar concepts, varying degrees of detail, and conflicting assumptions. There is no single, unifying frame of reference for validation. This has restricted the development of a common theoretical and practical understanding of validation across the M&S field. This paper introduces the concept of a validation frame of reference; builds a taxonomy of using frames of reference found in the validation literature; and proposes the creation of a unifying theoretical framework for M&S validation to build a foundation for the development of a paradigm of common theoretical and practical understanding of validation across the M&S field.

1   INTRODUCTION

Earliest discussions of validation are found in the fields of computer science, operations research, and mathematical modeling with practitioners drawing on quantitative and statistical validation techniques found in management science, behavior science, decision science, and discrete event simulation, to name a few (Sargent 1979). While many definitions for M&S validation abound, each generally contains the same core elements: the thing to be simulated, the simulation model, and a bounding principle (Moya and Youngblood 2007). The U.S. Department of the Defense (DOD) definition, derived from sources across the field of M&S, is a good example: "real world" describes the thing-to-be-simulated, "a model and its associated data" represents the simulation-model, and the combined concepts of "degree to which," "accurate representation," and "from the perspectives of the intended uses" define the bounding-principle.

Definition 1 Validation *is the process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.* (MSCO 2010)

The M&S field frequently links validation with verification. Verification addresses the computer code that representing a system model that ultimately will create the computer-based simulation – *was the model built correctly* (Sargent 1979). Validation focuses on the selection and development of the representation – *is this the correct model* (Sargent 1979). Verification deals with the transformational accuracy between models (Tolk 2012), and the determination of a code's correctness is well-studied in the field of computer science (i.a., Clarke and Wing 1996). Validation, the topic of this paper, deals with representational and behavioral accuracies (Tolk et al. 2011).

Despite the ubiquity of validation in the field of M&S, it has a myriad of connotations with overt and covert disagreements on its fundamentals, bearing striking resemblance to Kuhn's observations on the inherent differences in scientific understanding and practice between the social sciences and natural sciences (Kuhn 1970). To help develop clarity for the science in this vein, this paper gives a background overview of validation as it is approached in the field of M&S. As an organizing principle for understanding the works reviewed, it introduces the concept of a "validation frame of reference." Next, it identifies and

defines four of the primary frames-of-reference found in the validation literature today: procedures and their supporting validation standards, validity standards, frameworks, and theory. Finally, a solution to create a unified framework for validation is proposed.

2 BACKGROUND

Similar to the building of explanatory models in the realm of science, M&S validation has been regarded as a an activity governed by the scientific method. Within the field of practice, validators either seek evidence to falsify an M&S representation by analytically examining a model's formulation (*conceptual model validation*) and testing its simulation output (*results validation*) against system data (falsification) or assess and measure the match between system and simulation data (empiricism) (NGMS 2007; NGMS 2008; Heath and Hill 2010). While recognizing the influence of the scientific method on M&S validation practice, Heath and Hill (2010) observe that *no model* can be proven to be a valid representation. They use three different arguments to support their thesis: (1) reality is infinite with an infinite number of possible expressions and no model can represent an infinite system and can only (partially) represent one of those expressions of reality – there will always be a better model (cf., Heath and Hill 2010; Moya and Youngblood 2007; Robinson 1997); (2) Gödel's Incompleteness Theorem proves that for all but the simplest of systems, the set of axioms describing that system cannot be both consistent and complete – there will exist true statements about the system that cannot be expressed, or proven, using those axioms (Mendelson 2001); and (3) limitations of computing necessitate the incorporation factors not in the model but nevertheless required to make the model work (cf., Heath and Hill 2010; Winsberg 2010). While cogent and true, these arguments focus primarily upon the "accurate representation of the real world" portion of Definition 1. They neglect any qualifications or relaxations that might be allowed by the bounding principle: "from the perspective of the intended uses of the model."

Despite limitations on the ability to prove a model valid, or perhaps because of them, the DOD required assurance that using M&S, particularly for high-cost decisions or those potentially leading to high-consequence outcomes, would result in credible results. This led the DOD to develop the *Verification, Validation, and Accreditation Recommended Practices Guide* (*VV&A RPG*), now on its third edition (MSCO 2006). The *VV&A RPG* primarily acts as a procedural guide for practitioners across the model development and use cycle. It includes considerations for documentation and high-level descriptions of validation techniques and processes for various stages of simulation development. However, it does not provide a definitive guide for determining the techniques to be used in specific circumstances nor does it describe how to apply or implement the techniques (cf., Brade 2003, 32). Work has been done to codify these procedures into standards (e.g., IEEE SISC 2007; IEEE DISC 2010) and to collect best practices (e.g., Applegate, Blais, Burks, et al. 2011; Applegate, Blais, Brown, et al. 2011; Sanjay Jain and McLean 2009) for the community of practice. Petty, Dunning, and Collins (2011) have evaluated the community perception of effectiveness military M&S-standards, concluding that success is related to a standard's governance formality and technical specificity.

In addition to these efforts focused on the standard of validation practice, other efforts have focused on the standards for establishing model validity. Brade (2000; 2003) addressed limitations found in the *VV&A RPG* to establish methods and requirements for expressing the degree of validity achieved during a set of validation activities. His work aligns with fundamental precepts for confirming and building acceptability of scientific theories within the natural sciences, "The greater the variety [of quality tests having a favorable outcome], the stronger the support" (Hempel 1966, 34). The Validation Process Maturity Model (VPMM) also addresses the fact that validation assessments differ in their levels of objectivity, repeatability, timeliness, completeness, and accuracy and describes the elements to be included in a mature validation process (Harmon and Youngblood 2005). Its function is to provide a means to assess the validation process and the completeness of a validation assessment as proxies for the quality of the information it provides. It builds upon the Fidelity Framework developed by the Simulation Interoperability Standards Organization (SISO) to quantity simulation fidelity and facilitate the evaluation of a simulation. The developed Fidelity Framework considers model accuracy, error, fitness, precision, resolution, sensi-

tivity, tolerance, and validity as well as the concepts and roles of the "real world," application domain, fidelity referent, simulation requirements, acceptability criteria, and validation referent (Gross 1999). Similar to the procedural standards, these works provide little guidance on the selection of methods to achieve a desired degree of validity. To address the importance of understanding the degree of validation achieved, simulation credibility metrics and methods to measure against them have been proposed (Liu, Yang, and Wang 2005). Works focusing on the development of sound simulation-based studies also provide guidance on the selection of techniques and the establishment of criteria to measure model-validity. However, these generally focus on a specific modeling method (e.g., Law and McComas 2001; Law 2007; Bonté, Muller, and Duboz 2012) or are embedded within case studies (e.g., Bonté, Muller, and Duboz 2012; Tako and Robinson 2009) making generalization difficult.

Additional requirements necessary for model validity can be found in works not dedicated to the subject. These works relate to, having bearing on, or provide necessary considerations for validation and validity. Some of the more prevalent include discussions on data, accuracy, risk, and validity requirements. They describe criteria for measuring the validity of a model. For instance, the Levels of Conceptual Interoperability Model (LCIM) focuses on the requirements for model interoperability (Tolk, Diallo, and Turnista 2007). It implies necessity conditions for varying types or degrees of validity: data descriptions and format (syntactic), data meaning ranges (semantic), data use within the model (pragmatic), state trajectory transformation (dynamic), and model formulation (conceptual) (cf., Zeigler and Hammonds 2007; Tolk, Bair, and Diallo 2011). These levels are similar to the concepts of replicative validity, predictive validity, and structural validity found in the work of Zeigler, Praehofer, and Kim (2000).

Rather than focusing on the procedures, process, or standards of validation, some authors approach validation from its theoretical underpinnings. These works have borrowed from other fields upon which M&S depends to establish underlying principles specifically important to M&S. For instance, Discrete Event System Specification (DEVS) (Zeigler, Praehofer, and Kim 2000) and Model-Based Systems Engineering (MBSE) (Wymore 1993) draw from systems engineering, Model Based Data Engineering (MBDE) from database theory (Tolk and Diallo 2010), descriptions of M&S using model theory (Tolk et al. 2011), Simulation Composability Theory (SCT) from concurrency theory (Weisel, Petty, and Mielke 2003; Szabo, Teo, and See 2009), and Assumption Constraint Engineering (ACE) taking a computer science perspective (Tolk et al. 2011). From their foundational perspectives, these M&S theories develop principles, properties, definitions, conditions, theorems, and conjectures with respect to M&S in general, some of which affect the practice, assessments, and limits of validation in particular. For instance, MBDE establishes data dependencies for valid interoperable models; SCT gives results for the validity of the composition of valid models when the composition exists, proving that the composition of valid models is not necessarily valid; and ACE demonstrates that even models that are aligned from a data perspective can result in an invalid composition if using different underlying assumptions. While the expressiveness and power of DEVS, MBSE, and SCT have been mapped and compared (Mielke, Petty, and Weisel 2005; Weisel, Petty, and Mielke 2005), the other theoretical foundations have not been aligned.

Other works approach validation conceptually by examining the relationships between the artifacts created during M&S development and the thing or system represented by the M&S. The procedures, standards, and theories already discussed generally have a conceptual framework that guided their development. These may be explicitly described (cf., the Problem Solving Process in the *VV&A RPG*, Brade's V&V Phases and Sub-phases; the REVVA Generic Process developed for the West European Armament Group (Jacquart 2004)) or implicitly inferred (cf., the concepts of Natural System and comparisons between labeled transition systems in SCT; experimental frame implications in DEVS). In fact, the Sargent Circle of Validation (Sargent 1979; Sargent 2007) was developed initially to support the American Society of Mechanical Engineers in developing a verification and validation guide for computational solid mechanics (Pace 2004).

While there are several of these conceptual frameworks for validation, they generally align with the model development process and there are many commonalities between them. Bair (2012) discusses many of these common elements within these frameworks which include application objectives, thing being modeled, conceptual model, simulation results, simulation specification, and simulation model (cf.,

MSCO 2006; MSG-054/TG-037 2012; Sargent 2007; Weisel 2011; Balci and Ormsby 2007). The primary differences between the frameworks lie in level of granularity and terminology. For instance, some authors focus on conceptual model validation (e.g., Nance 1984; Balci and Ormsby 2007; Robinson et al. 2011). On the other hand, Weisel (2011; 2012) focuses the relationships between validation theory, abstract constructs, and practical applications creating linkages between findings found computational science, decision theory, and M&S. The Sargent Circle of Validation focuses on artifacts available during the modeling development process as they relate to verification and validation activities (Sargent 2007). While Brade (2003; 2000) adds granularity by explicitly tying these artifacts to the model-development process and specific questions to be asked and comparisons to be made during it.

3 VALIDATION FRAMES OF REFERENCE

The use and application of the scientific method to test scientific theory is intrinsically tied to the subject matter within the theory's field of science itself (Radner and Radner 1982).

> In the absence of a paradigm or some candidate for paradigm, all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant. As a result, early fact-gathering is a far more nearly random activity than the one that subsequent scientific development makes familiar. Furthermore, in the absence of a reason for seeking some particular form of more recondite information, early fact-gathering is usually restricted to the wealth of data that lie ready to hand. The resulting pool of facts contains those accessible to casual observation and experiment together with some of the more esoteric data retrievable from established crafts like medicine, calendar making, and metallurgy. (Kuhn 1970, 15)

The paradigm used to guide scientific thought within a field determines its legitimate problems, admissible solutions, and allowable techniques. Kuhn observes that a mature scientific community will operate within a single or small set of paradigms, and if multiple scientific communities investigate the same problems, then there generally is significant overlap of their guiding paradigms. The paradigm of a field of study expresses the consensus among its practitioners on the field's important and fundamental findings. These principles are no less true for the practice of M&S validation within the field of M&S. Kuhn remarks, "Despite occasional ambiguities, the paradigms of a mature scientific community can be determined with relative ease" (Kuhn 1970, 43). However, there are indications that the field of M&S, particularly with respect to validation lacks this desired maturity: "If in doubt collect [the data for model validation]" (Cowdale 2009, 1768). It, therefore, is worth asking: What is, or are, the paradigms governing M&S validation?

Mirriam-Webster Online defines paradigm consistent with the ideas expressed Kuhn:

Definition 2 *A* paradigm *is a philosophical and theoretical framework of a scientific school or discipline within which theories, laws, and generalizations and the experiments performed in support of them are formulated.* (m-w.com 2013)

A taxonomy that structures the findings and guiding principles within the body of M&S literature will help realize the stringent burden imposed by this definition, namely the M&S governing philosophy, theories, laws, generalizations, and allowable experiments. The concept of a frame of reference will aid understanding the types of findings within the body of literature and their classification.

Definition 3 *A* frame of reference *is a set of ideas, conditions, or assumptions that determine how something will be approached, perceived, or understood.* (m-w.com 2013)

A frame of reference is less strict than that of paradigm, although in a mature field of study all frames of reference would fit within the prevailing paradigm. Rather, one can think of a paradigm as consolidating into consensus the various lenses through which the various practitioners view validation. Classifying the findings resident within these works can provide the basis for the uncovered or to-be developed M&S validation paradigm that will, eventually, reach the ideals set by the Definition 2. The developed taxonomy proposes four frames-of-reference: procedural guidelines with supporting validation standards

(Procedures and Validation Standards), standards for ensuring model validity (Validity Standards), a conceptual framework (Frameworks), or supporting theory (Theories).

3.1   Procedures and Validation Standards

Procedures are generally concerned with developing and providing documentation of the model, simulation, code, testing procedures, and testing results.

Definition 4 *A* procedure *is an established or particular way of accomplishing something, often expressed as a series of steps to be followed in a regular definite order.* (m-w.com 2013)

Available procedures encompass methods, tools, checklists, steps, templates for documentation and similar, and the community accepted and expected practices reside in validation standards.

As observed by Petty, Dunning, and Collins (2011) there is significant commonality between that which is described as procedures here and by standards within the community of practice, which this paper terms *validation standards*. To clarify the difference between validation standards and validity standards as used in this paper, consider Definition 5.

Definition 5 *A* standard *is (a) something set up and established by authority as a rule for the measure of quantity, weight, extent, value, or quality; (b) something established by authority, custom, or general consent as a model or example.* (m-w.com 2013)

Within the taxonomy created here, a *validation standard* adopts meaning (b) in Definition 5. The meaning in (a) is reserved for what is termed *validity standard*.

Works within this frame of reference generally consider the steps of model-building but in terms of the validation activities required. They concentrate on the activities required to collect information to be provided to decision-makers and the documentation necessary to support their assessments. Validation results focus on generating sufficient information, reporting, and documentation across all phases of the model develop cycle, or alternatively the artifacts generated during model development, to generate model-credibility for the user. Examples include the *VV&A RPG* and *IEEE Std 1516.4™-2007*. Each describes validation activities required, participants, documentation requirements, etc. In this respect, they serve as community-accepted models for performing validation. While the *VV&A RPG* includes a comprehensive collection of techniques, it has no guidance for application or interpretation of results. Rather, their general focus is on process, implying that the application of a good process with satisfactory findings will equate to valid models. Unfortunately, the procedures provided are ill-defined allowing for significant interpretation. Brade attributes this to its being a political document (Brade 2003, 32).

There is significant overlap between the communities developing and supporting the *VV&A RPG* and *IEEE Std 1516.4™-2007*. Law and McComas (2001) provide an alternative viewpoint from the field of analysis. They give more emphasis on the application and selection of techniques, but the collection of choices is far fewer. Guidance on which data are important is lacking, primarily due to the strong dependency on the problem of interest. Finally, the focus on solving an analytical problem drives the focus of the validation activities described on specifically evaluating the ability to solve the problem(s) posed rather than a more general validation presumed by the other procedural examples.

Ideally, procedural approaches to validation, bolstered by community supported standards, would give the practitioner a set of steps to follow, methods to draw upon, and tools to consider when embarking on a validation study. Current, procedural approaches, even supported by community-accepted standards, fall short of this ideal, lacking detailed step-by-step guidance. In terms of fact-gathering discussed by Kuhn, which is the primary role of all validation activities, this lack of guidance makes data collection far more random than one would hope. This may be another reason underlying Heath and Hill's conjecture that "simulation validation in practice is really the process of persuading the evaluators to believe that the simulation is valid with respect to the objective" (Heath and Hill 2010, 31).

3.2   Validity Standards

The concept of validity standard used in this paper follows the sense of (a) in Definition 5, differing from the use found in Petty, Dunning, and Collins (2011). A validity standard, as used here, addresses the

measure of quality achieved by a set of validation activities. These works eschew the assumption that good process with satisfactory findings will equate to valid models. Rather, these standards attempt to address the quality of information generated by those activities irrespective of the process followed. They are outcome oriented and provide measurement criteria and necessity conditions for establishing validity.

While lacking strong governance authority, the VPMM (Harmon and Youngblood 2005) and Fidelity Framework (Gross 1999) more precisely addresses measuring model validity than that found in more formal validation standards. The VPMM suggests three measures of information quality derived from a validation process: completeness, correctness, and confidence. Liu, Yang, and Wang (2005) suggest that in addition to measuring the validity of an M&S, other metrics of a model's credibility are important: reliability, usability, correctness, and interoperability. Correctness in the latter view reflects the correctness of the computer code used to generate the simulation determined via verification, while correctness in the former sense reflects a measurement of the error between the system being modeled and the simulation-model representation of that system.

Harmon and Youngblood identify several elements that ideally should be included in any validation assessment. These can be broken into two main categories: the validation criteria and an assessment of the correspondence between the requirements found in the validation criteria and the simulation. The validation criteria used in the assessment includes descriptions of the entities, properties, and dependencies required to meet the user's purpose; acceptable error tolerances for a given set of input ranges; and required confidence values. Harmon and Youngblood also propose six tiers of validity assessment based on the type of information collected and the resulting validity statements that can be made as a result. These tiers range from "trust me" to statistical measures of confidence, and primarily address their proposed measures of correctness and confidence. This begins to provide a guide to practitioners on the types of data they need to collect and the tests they need to run to achieve the desired outcome rigor from their validation efforts. The work of Liu, Yang, and Wang supports this general approach.

The Fidelity Framework provides metrics to assess the match between a simulation-model and the system it represents. These provide considerations when collecting data and selecting validation tests. However, the validator still needs to decide which system elements and data are important. That is, in part, governed by the framework in which the validation problem is viewed: e.g., a faithful match of reality (Heath and Hill 2010), sufficient to answer the analytic problem of interest (Law and McComas 2001), or able to support experimentation with and further understanding of the system modeled (Sargent 2007).

3.3 Frameworks

Validation frameworks available within the M&S literature address the intersection between M&S development and its testing. They identity the constituent elements of validation and the relationships between those elements. These validation frameworks drive the conceptual thought for practitioners as they develop and apply procedures, methods and tools to validation and for theorists as they develop the theoretical foundations for modeling and simulation. These thoughts lead to Definition 6.

Definition 6 *A* framework *is a basic conceptual structure; that is, an organizational aggregate of the basic elements of an entity in their relationships to each other as abstracted or generalized from particular instances and dominated by the general character of the whole.* (m-w.com 2013)

To see how frameworks influence validation practice, consider the Sargent Circle (Sargent 2007) and the Simulation Conceptual Modeling Lifecycle (Balci and Ormsby 2007). When comparing the two frameworks, three observations are immediately obvious: both address conceptual model validation, only the Sargent Circle addresses results validation, only the Simulation Conceptual Modeling Lifecycle addresses data validation. Further, irrespective of the differences in terminology, the granularity in the two frameworks in representing the relationships with respect to the conceptual model, the system of interest, and the simulation-model differ. The first and last of these is understandable given the conceptual modeling focus of the second framework. In each of these frameworks, practitioners can reconcile the set of activities required, the comparisons to be made under test, and the artifacts to consider. The differences between the two frameworks in their representations of the validation and modeling activities with respect

to the system of interest potentially reflect paradigm differences. The Sargent Circle reflects a paradigm driven by scientific exploration, theorizing, and testing. By contrast, the Simulation Conceptual Modeling Lifecycle reflects a paradigm driven by problem solving and analysis. When applying either of these frameworks or another framework found in the field of practice, the selection of framework drives the elements important to evaluate, the types of evaluations made, and the types of questions asked.

3.4   Theories

Theories for M&S in general and applied to M&S validation in particular look to extend the science of M&S using the language of mathematics. They examine its foundational elements and fundamental problems and identify its underlying assumptions. Theories in this sense rely on formal methods. *Formal* methods are mathematically based techniques that allow the use of mathematical representations and logic based proofs in the descriptions of simulation models and their resultant simulated behavior. They are in contrast to formalized methods that use various notations, diagrams, and descriptions to systematically describe the requirements for a model but still can result in reduce ambiguity, imprecision, and misunderstandings or necessitate premature decisions in design due to the nature of the formalized method of description selected (Rushby 1993).

Theoretical literature includes the development of formal frameworks for M&S related concepts and issues that allow reasoning upon their structures. They further allow the development of properties and theorems of M&S based upon those structures described by the formal theory to further understand the problems within the field of study. Within M&S, they can be applied to M&S issues related to interoperability, data, validation, and others. The conceptual frameworks discussed in the previous section provide motivating considerations for what is known, unknown, and important to understand within M&S validation. This gives rise to Definition 7.

Definition 7 *A* theory *is an analysis of a set of facts in their relation to each other that provides a plausible or scientifically acceptable general principle or body of principles offered to explain phenomena described by a body of theorems presenting a concise systematic view of a subject.* (m-w.com 2013)

While there is work on the process of model building, and the field borrows from computational science for the underpinnings of simulation development, there is little theory developed specific to M&S. It lacks a robust foundational theory that contains formal definitions and theorems that prove boundaries of use, limits of validity, necessary and sufficient properties for interoperability, and more. Although there are few theoretical works available in the M&S body of literature, the number is not zero. DEVS is perhaps the most well-known (Zeigler, Praehofer, and Kim 2000), with SCT providing a lesser-known alternative view (Weisel, Petty, and Mielke 2003).

SCT focuses on the validity of component based simulation. It provides a theoretical foundation, based in computational computer science, to determine the validity of model composed of two or more other models. Findings from this work show that in general the component selection problem is NP-complete and demonstrated that the validity of a composed model should never be taken for granted, even if each of the models within the composition are themselves valid. MBDE gives additional theoretical findings: e.g., there is interoperability between models if and only if there is a common reference model determined from the common intersection between their existential and transformational dependencies (Tolk and Diallo 2010).

Comparisons between SCT and DEVS showed that both have the same computational power for computer implemented simulation-models (Weisel, Petty, and Mielke 2005). This comparison concludes that while DEVS has advantages for system design, SCT has advantages for the theoretical investigations. These differences clearly reflect the differences in framework applied to their theory development. DEVS, as a systems-engineering based approach, primarily concerns itself with specifying models. This is explicit in its name, and a cursory review of the literature supports this conclusion. In contrast, the purpose SCT is to enable the development of new theories and build understand about simulation and simulation-validation. The work of Szabo, Teo and See (2009) supports this purpose.

These works, and others like it, provide inroads for building a robust theory of M&S with insights for validation. However, this field lies much in the same state as the fledgling beginning to calculus Newton and Leibniz argued over notation and who invented the theories first. Nor does it differ much from the beginnings of computational theory with the differing structures found in lambda calculus and Turing machines. The lack of a common notational structure prevents meaningful discourse and strides of progress in understanding the underlying theory to M&S. This inherently reflects the lack of common paradigm.

4 CONCLUSIONS AND WAY AHEAD

Despite all of the existing validation literature, work continues because the questions plaguing our understanding validity are the same questions at the heart of any scientific endeavor. When reviewing the body of literature, it becomes immediately evident that the various authors primarily approach their discussions using one of four viewpoints, or frames of reference: procedural guidelines with supporting validation standards (Procedures and Validation Standards), standards for ensuring model validity (Validity Standards), a conceptual framework (Frameworks), or supporting theory (Theories). All of these works provide frames of reference for understanding and applying M&S validation. However, despite the breadth covered by this validation literature, the works are disjoint. For example, common terms are used differently; differing terms are used to refer to similar concepts; some concepts exist within some works but not in others; and varying degrees of specificity and granularity exist.

Further, none of the frames of reference support all aspects: procedures, standards, theory and guiding framework for thought; although, few works could be considered solely or completely within one of these frames of reference. Most works include findings or, at least, assumptions across more than one of these taxonomic categorizations. However, none of these works span all of these needs. Rather, they generally focus the work through one of these lenses with information relevant to the other categories merely a derivative of or supporting element to their discussion. In fact, it is fair to say that procedural standards tell the practitioner *what to do and how to do it*, validity standards *ensure that those activities meet the objectives*, conceptual frameworks help *ensure their consistency*, and theories *explain why* the procedures and conditions set by the standards work (Figure 1).
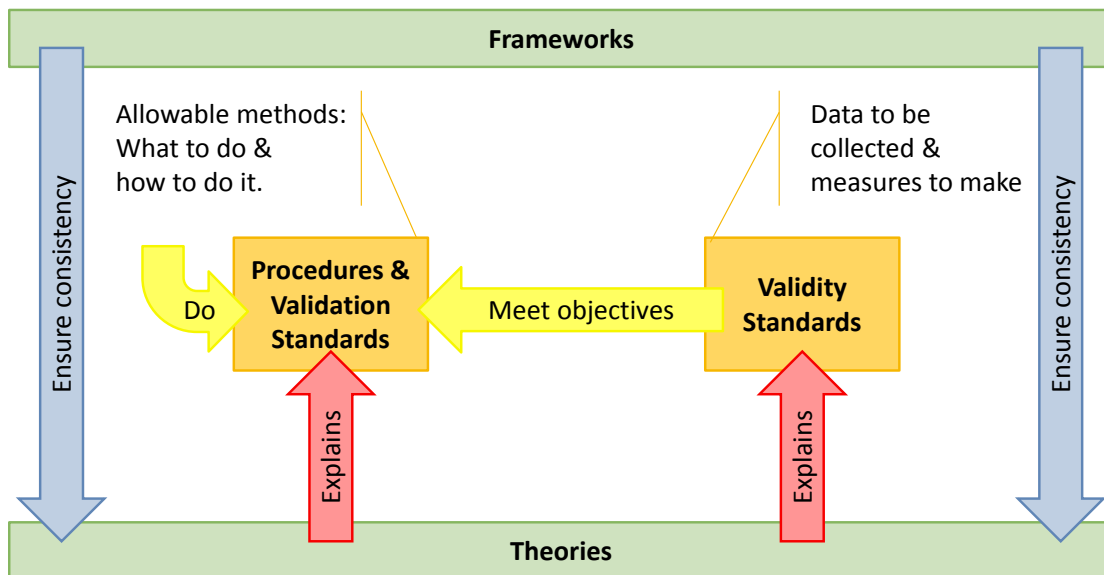

Figure 1: M&S validation frames-of-reference taxonomy and category relationships

This taxonomy shows that there is no single, unifying frame of reference for modeling and simulation validation. The lack of alignment and unified thought has led to inconsistencies across the body of work and reflects the lack of overarching M&S validation paradigm in the sense discussed by Kuhn. This has

resulted in the development of multiple validation frameworks, procedures, and standards that are unaligned and inconsistent with each other. It has also restricted the development of a common theoretical and practical understanding of validation across the modeling and simulation field.

This work represents the first step in ongoing research currently being conducted. The work of Tolk, Padilla, and Diallo (2012) suggests that more formal approaches to validation could help provide the needed common language. Formal methods give an underlying mathematical basis to validation that ensures clarity, creates specificity, enables repeatability, and provides access to unambiguous and traceable theorem-provers. The next steps in this research will include evaluating the terms used across the M&S validation literature on the road to formally defining and aligning validation terms.

Irrespective of the approach taken, a unifying frame of reference for M&S validation is necessary to build a foundation for the development of paradigm for a common theoretical and practical understanding of validation – its frameworks, procedures, standards, and theories – across the M&S field.

REFERENCES

Applegate, J., C. Blais, R.F. Brown, M. Jaye, and E.W. Weisel. 2011. "Irregular Warfare (IW) Data Quality Best Practices Guide". Technical Report. Monterey, CA: TRADOC Analysis Center.

Applegate, J., C. Blais, R. Burks, R.F. Brown, D. Duong, M. Jaye, T. Perkins, and M. Thompson. 2011. "Irregular Warfare (IW) Model Validation Best Practices Guide". Technical Report. Monterey, CA: TRADOC Analysis Center.

Bair, L.J. 2012. "The Missing Link in Modeling and Simulation Validation." In *Proceedings of the 2012 Autumn Simulation Conference*. San Diego, CA: Society for Modeling and Simulation International (SCS). http://scs.org/autumnsim/autumnsim2012/papers.

Balci, O., and W.F. Ormsby. 2007. "Conceptual Modelling for Designing Large-scale Simulations." *Journal of Simulation* 1 (3) (August): 175–186.

Bonté, B., J-P. Muller, and R. Duboz. 2012. "NATO MSG-088 Case Study Results to Demonstrate the Benefit of Using Data Farming for Military Decision Support." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher. Piscataway, NJ: IEEE.

Brade, D. 2000. "Enhancing Modeling and Simulation Accreditation by Structuring Verification and Validation Results." In *Proceedings of the 2000 Winter Simulation Conference*, edited by J.A. Joines, R.R. Barton, K. Kang, and Paul. A. Fishwick, 840–848. Piscataway, NJ: IEEE.

Brade, D. 2003. "A Generalized Process for the Verification and Validation of Models and Simulation Results". Doctor rerum naturalium, Neubiberg, Germany: Fakultät für Informatik Universität der Bundeswehr München.

Clarke, E.M., and J.M. Wing. 1996. "Formal Methods: State of the Art and Future Directions." *ACM Computing Surveys* 28 (4) (December 1): 626–643.

Cowdale, A. 2009. "Real World Data Collection for Model Validation." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M.D. Rosetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, 1764–1769. Piscataway, NJ: IEEE.

Gross, D.C., ed. 1999. "Report from the Fidelity Implementation Study Group." In *Proceedings of the Spring 1999 Simulation Interoperability Workshop*, 99S–SIW–167 (SISO–REF–02–1999). Orlando, FL: Simulation Interoperability Standards Organization (SISO). http://www.sisostds.org/.

Harmon, S.Y., and S.A. Youngblood. 2005. "A Proposed Model for Simulation Validation Process Maturity." *Journal of Defense Modeling and Simulation* 2 (4) (October): 179–190.

Heath, B.L., and R.R. Hill. 2010. "Final Report Developing an Agent-based Modeling Verification and Validation Approach for Improving Air Force Analytical Support". Dayton, OH: BIE Department, Wright State University. http://orsagouge.pbwiki.com/ABSVal.

Hempel, C.G. 1966. *Philosophy of Natural Science*. Englewood Cliffs, N.J.: Prentice-Hall.

IEEE Computer Society. Distributed Interactive Simulation Committee, Institute of Electrical and Electronics Engineers, and IEEE-SA Standards Board. 2010. *IEEE Recommended Practice for Distributed Interactive Simulation Verification, Validation, and Accreditation*. New York: IEEE.

IEEE Computer Society. Simulation Interoperability Standards Committee, Institute of Electrical and Electronics Engineers, and IEEE-SA Standards Board. 2007. *IEEE Recommended Practice for Verification, Validation, and Accreditation of a Federation an Overlay to the High Level Architecture Federation Development and Execution Process*. New York: IEEE.

IEEE DISC. See also IEEE Computer Society. Distributed Interactive Simulation Committee.

IEEE SISC. See also IEEE Computer Society. Simulation Interoperability Standards Committee.

Jacquart, R. 2004. "VV&A Methodological Guidelines Reference Manual (METHGU2)". THALES JP 11.20-WE5100–METHGU2-D5103. Technological Arrangement for Laboratories Defence European Studies Joint Program 11.20. http://www.revva.eu.

Jain, S, and C.R. McLean. 2009. "Recommended Practices for Homeland Security Modeling and Simulation." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M.D. Rosetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, 2879–2890. Piscataway, NJ: IEEE.

Kuhn, T.S. 1970. *The Structure of Scientific Revolutions. Second Edition, Enlarged.* Chicago, London: University of Chicago Press. http://turkpsikiyatri.org/arsiv/kuhn-ssr-2nded.pdf.

Law, A.M. 2007. *Simulation Modeling and Analysis*. 4th ed. Boston: McGraw-Hill.

Law, A.M., and M.M. McComas. 2001. "How to Build Valid and Credible Simulation Models." In *Proceedings of the 2001 Winter Simulation Conference*, edited by B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer, 22–29. Piscataway, NJ: IEEE.

Liu, F., M. Yang, and Z. Wang. 2005. "Study on Simulation Credibility Metrics." In *Proceedings of the 2005 Winter Simulation Conference*. Piscataway, NJ: IEEE.

m-w.com. 2013. *Mirriam-Webster Online*. Springfield, MA: Mirriam-Webster Incorporated. http://www.merriam-webster.com/.

Mendelson, E. 2001. *Introduction to Mathematical Logic*. London: Chapman & Hall.

Mielke, R.R., M.P. Petty, and E.W. Weisel. 2005. "A Comparison of MBSE and Semantic Composability Theory." In *Proceedings of the 2005 Huntsville Simulation Conference*. Vista, CA: The Society for Modeling and Simulation International.

Modeling and Simulation Coordination Office. 2006. *Verification, Validation, and Accreditation Recommended Practices Guide*. RPG Build 3.0: September 2006. Alexandria, VA: Department of Defense. http://vva.msco.mil/.

Modeling and Simulation Coordination Office. 2010. "DoD Modeling and Simulation (M&S) Glossary". Final Report for High Level Task S-C-3. Alexandria, VA: Department of Defense. http://www.msco.mil/MSGlossary.html.

Modeling and Simulation Group 054 / Task Group 037. 2012. "Risk-based Tailoring of the Verification, Validation, and Accreditation / Acceptance Processes". RTO-TR-MSG-054. Brussels, Belgium: North Atlantic Treaty Organisation, Research and Technology Organisation, panel NMSG. http://www.cso.nato.int/abstracts.aspx.

Moya, Lisa Jean. See also Bair, Lisa Jean.

Moya, L.J., and S.A. Youngblood. 2007. "Clarifying Validation for Agent Based Simulations." In *Proceedings of the 2007 Fall Simulation Interoperability Workshop*, 07F–SIW–104. Orlando, FL: Simulation Interoperability Standards Organization (SISO). http://www.sisostds.org/.

MSCO. See also Modeling and Simulation Coordination Office.

MSG-054/TG-037. See also Modeling and Simulation Group 054 / Task Group 037.

Nance, R.E. 1984. "Model Development Revisited." In *Proceedings of the 1984 Winter Simulation Conference*, 75–80. Piscataway, NJ: IEEE.

NGMS. See also Northrop Grumman Mission Systems.

Northrop Grumman Mission Systems. 2007. "The Agent-based Simulation (ABS) Verification, Validation, and Accreditation (VV&A) Study Final Report". M00264-06-D-0001-DO005. Quantico, VA:

Marine Corps Combat Development Command (MCCDC) Operations Analysis Division (OAD). http://orsagouge.pbwiki.com/ABSVal.

Northrop Grumman Mission Systems. 2008. "The Agent-based Simulation (ABS) Verification, Validation, and Accreditation (VV&A) Study Phase II - Joint/DOD Final Report". M00264-06-D-0001-DO005. Quantico, VA: Marine Corps Combat Development Command (MCCDC) Operations Analysis Division (OAD). http://orsagouge.pbwiki.com/ABSVal.

Pace, D.K. 2004. "Modeling and Simulation Verification and Validation Challenges." *Johns Hopkins APL Technical Digest* 25 (2): 163–172.

Petty, M.P., R.L. Dunning, and A.J. Collins. 2011. "Correlation of Characterizing Attributes and Success in Military M&S Standards." In *Proceedings of the Fall 2011 Simulation Interoperability Workshop*, 11F–SIW–027. Orlando, FL: Simulation Interoperability Standards Organization (SISO). http://www.sisostds.org/.

Radner, D., and M. Radner. 1982. *Science and Unreason*. Belmont, Calif.: Wadsworth.

Robinson, S. 1997. "Simulation Model Verification and Validation." In *Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson, 53–59. Piscataway, NJ: IEEE.

Robinson, S., R. Brooks, K. Kotiadis, and D-J. van der Zee, ed. 2011. *Conceptual Modeling for Discrete-event Simulation*. Boca Raton: CRC Press.

Rushby, J. 1993. "Formal Methods and the Certification of Critical Systems". Technical Report CSL-93-7. Menlo Park, CA: Computer Science Laboratory SRI International.

Sargent, R.G. 1979. "Validation of Simulation Models." In *Proceedings of the 1979 Winter Simulation Conference*, edited by Mitchell G. Spiegel, Robert Shannon, and Harold J. Highland, 497–503. San Diego, CA: IEEE.

Sargent, R.G. 2007. "Verification and Validation of Simulation Models." In *Proceedings of the 2007 Winter Simulation Conference*, edited by S.G. Henderson, B. Biller, M.-H. Hsieh, J.D. Shortle, J.D. Tew, and R.R. Barton, 124–137. Piscataway, NJ: Institute of Electrical and Electronics Engineers. http://informs-sim.org/.

Szabo, C., Y.M. Teo, and S. See. 2009. "A Time-based Formalism for the Validation of Semantic Composability." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M.D. Rosetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, 1411–1422. Piscataway, NJ: IEEE.

Tako, A.A., and S. Robinson. 2009. "Comparing Model Development in Discrete Event Simulation and System Dynamics." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M.D. Rosetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, 979–991. Piscataway, NJ: IEEE.

Tolk, A. 2012. "14: Verification and Validation." In *Engineering Principles of Combat Modeling and Distributed Simulation*, edited by A. Tolk. Hoboken: Wiley.

Tolk, A., L.J. Bair, and S.Y. Diallo. 2011. "Supporting Network Enabled Capability by Extending the Levels of Conceptual Interoperability Model to an Interoperability Maturity Model." *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 10 (2) (December 6): 145–160.

Tolk, A., and S.Y. Diallo. 2010. "Using a Formal Approach to Simulation Interoperability to Specify Languages for Ambassador Agents." In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 359–370. Piscataway: IEEE.

Tolk, A., S.Y. Diallo, H. Herencia-Zapana, J.J. Padilla, R.R. Creasey, J. Himmelspach, M.P. White, and M. Fu. 2011. "Model Theoretic Implications for Agent Languages in Support of Interoperability and Composability." In *Proceedings of the 2011 Winter Simulation Conference*, edited by S Jain, 309–320. Piscataway: IEEE.

Tolk, A., S.Y. Diallo, R.D. King, C.D. Turnista, and J.J. Padilla. 2011. "14. Conceptual Modeling for Composition of Model-based Complex Systems." In *Conceptual Modeling for Discrete-event Simulation*, edited by S. Robinson, R. Brooks, K. Kotiadis, and D-J. van der Zee. Boca Raton: CRC Press.

Tolk, A., S.Y. Diallo, and J.J. Padilla. 2012. "Semiotics, Entropy, and Interoperability of Simulation Systems – Mathematical Foundations of M&S Standardization." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher. Piscataway, NJ: IEEE.

Tolk, A., S.Y. Diallo, and C.D. Turnista. 2007. "Applying the Levels of Conceptual Interoperability Model in Support of Integratability, Interoperability, and Composability for System-of-Systems Engineering." *Journal of Systemics, Cybernetics and Informatics* 5 (5): 65–74.

Weisel, E.W. 2011. "Towards a Foundational Theory for Validation of Models and Simulations." In *Proceedings of the Spring 2011 Simulation Interoperability Workshop*, 11S–SIW–074. Orlando, FL: Simulation Interoperability Standards Organization (SISO). http://www.sisostds.org/.

Weisel, E.W. 2012. "Decision Theoretic Approach to Defining Use for Computer Simulation." In *Proceedings of the 2012 Autumn Simulation Conference*. San Diego, CA: Society for Modeling and Simulation International (SCS).

Weisel, E.W., M.P. Petty, and R.R. Mielke. 2003. "Validity of Models and Classes of Models in Semantic Composability." In *Proceedings of the Fall 2003 Simulation Interoperability Workshop*, 03F–SIW–073. Orlando, FL: Simulation Interoperability Standards Organization (SISO). http://www.sisostds.org/.

Weisel, E.W., M.P. Petty, and R.R. Mielke. 2005. "A Comparison of DEVS and Semantic Composability Theory." In *Proceedings of the Spring 2005 Simulation Interoperability Workshop*, 956 – 964 (05S–SIW–153). San Diego, CA: Simulation Interoperability Standards Organization (SISO). http://www.sisostds.org/.

Winsberg, E.B. 2010. "Sanctioning Models: Theories and Their Scope." In *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.

Wymore, A.W. 1993. *Model-based Systems Engineering: An Introduction to the Mathematical Theory of Discrete Systems and to the Tricotyledon Theory of System Design*. Boca Raton: CRC Press.

Zeigler, B.P., and P. Hammonds. 2007. *Modeling & Simulation-based Data Engineering Introducing Pragmatics into Ontologies for Net-centric Information Exchange*. Burlington, MA: Elsevier.

Zeigler, B.P., H. Praehofer, and T.G. Kim. 2000. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. San Diego: Academic Press.

AUTHOR BIOGRAPHIES

LISA JEAN BAIR is founder and principal of Lisa Jean Bair Analytics LLC. Ms. Bair is a Ph.D. candidate in the M&S program at Old Dominion University (ODU) and earned an M.S. in Operations Research from The College of William and Mary and a B.S. in Applied Mathematics from ODU. Her email address is lisa@lisajeanbair.com.

ANDREAS TOLK is Chief Scientist for SimIS Inc. in Portsmouth Virginia USA where he is responsible for implementing long term visions regarding developments in simulation. He is also Adjunct Professor at Old Dominion University in Norfolk Virginia USA. He holds a M.S. and Ph.D. in Computer Science from the University of the Federal Armed Forces in Munich, Germany. His email address is andreas.tolk@simisinc.com.