The Validity Tests Used by
Social Scientists and Decision Makers

Adam Eckerd
School of Public Affairs and Administration
University of Kansas
4060 Wescoe Hall
1445 Jayhawk Blvd.
Lawrence, Kansas 66045
Phone: (785) 864-3527
aeckerd@ku.edu

David Landsbergen
Anand Desai
John Glenn School of Public Affairs
110 Page Hall
The Ohio State University
1810 College Road
Columbus, Ohio 43210

Phone: (614) 292-9577
landsbergen.1@osu.edu

Phone: (614) 292-0826
desai.1@osu.edu

How can simulation be "sold" to policy decision makers? How can simulation be sold to other social scientists that do not accept simulation as a complement to "accepted" techniques" (Repenning, 2003)? Decision makers and social scientists use validation tests to determine how much confidence they should vest in a model (Forrester and Senge, 1980). And because these communities have different uses for models, they will employ different validation tests. If validation tests are not sufficiently clear, several problems could occur. A decision-maker may "dismiss" a simulation model using a particular validation test, unbeknownst to the modeler. This paper collects the validation tests in the various simulation and statistical / psychometric literatures into a comprehensive framework. Decision-makers are keenly interested in use as well as how "scientifically valid" that model is. Therefore, there are analytic, "consequential", and pragmatic validity tests. Decision-makers may rely on heuristics (Kahneman, Slovic and Tversky, 1982) as validity tests because they have difficulty understanding simulation (Cronin, Gonzalez, & Sterman, 2009). Decision-makers can teach modelers too – humans have used heuristics to become the dominant species in complex ecosystems. The long-term research objective is to use experiments (Sterman, 1987) to understand how decision-makers use both adaptive and dysfunctional heuristics.

The Validity Tests Used by Social Scientists and Decision Makers

Introduction

How can simulation be "sold" to policy decision makers? How can it be sold to other social scientists that do not understand or accept simulation as a valid complement to linear, statistical, or econometric models (Repenning, 2003)? We propose that the discourse surrounding the "old questions" in epistemology and ontology have much to contribute.

This appeal to philosophy has a very practical focus – the use of validation tests to increase use and understanding. Both decision makers and social scientists use validation tests to determine how much confidence they should vest in a model (Forrester and Senge, 1980). And because these communities have different uses for models, they will employ different validation tests.

If these validation tests are not sufficiently clear, explicit, or not fully identified, several problems could occur. For example, a decision-maker may judge the value of a simulation model using a validation test, unbeknownst to the modeler. The first contribution of this line of research, therefore, is to collect many of the validation tests cited in the simulation and statistical / psychometric literatures. These validation tests are collected together into a process framework by which a simulation model would be first proposed, built, and then utilized (Yucel and van Daalen, 2009).

Having collected these validation tests in one place, the validation tests can then be used to examine the "marketing question." Policy decision-makers are keenly interested in use as well as how "scientifically valid" that model is. Therefore, there is both scientific and pragmatic utility, and consequently, scientific and pragmatic validity tests.

This means that not only must the model help the policy decision-maker with their particular problem, but it also must appeal to the way in which a decision-maker perceives, understands, and judges the knowledge that is generated by a simulation model. There is no compromise because of the validity tests go beyond scientific validity, since the decision-maker must understand and accept the knowledge if they are going to use it. One of the long-term research objectives is to build upon the existing literature on the use of simulations by exploring how decision-makers employ the various validation tests, or "heuristics", to judge the usefulness of simulation models.

This effort is not a crass exploitation that one might see in the use of cognitive and behavioral psychology by a marketing firm. It is merely communicating the knowledge in a way that makes sense to decision-makers.

In some cases, decision-makers may have significant difficulty understanding some of the core ideas in simulation (e.g. systems dynamics (Cronin, Gonzalez, & Sterman, 2009)). But it also may be the case that decision-makers have much to tell modelers about how the human brain and human species has managed to become (unfortunately) the (too) dominant species in complex ecosystems. Much like industrial designers look to shark-skin to design boat hulls so that they have less drag or look to botany for inspiration in the design of Velcro, there may be adaptations in the human brain that allows humans to simplify the problem of surviving in complex environments in ways that may help build simulation models. Experiments might be very useful in helping test simulation models to identify both adaptive as well as dysfunctional decision heuristics (Sterman, 1987).

Kuhn's (1962) *Structure of Scientific Revolutions* is a strong argument for the social and cultural determinants of accepted scientific knowledge. Drawing upon the work of Kahneman, Slovic, and Tversky (1982) it is clear that scientists also use heuristics in their construction and testing of scientific knowledge (Faust, 1982). Clearly, just as it is important to identify the particular validation tests used by decision-makers, it is also important to identify the local validation tests used by social scientists as well.
Even if do not accept the social construction view of the generation of scientific knowledge, identifying the various validation tests used in communities of scientists can still provide a common ground with mainstream modelers who make decisions about who is doing interesting research. If shared tests for quality work can be found, it may even lead to research models which combine statistical / econometric approaches with simulation.

Finally, gathering together the variety of validation tests identified in the literature in one place, might provoke a closer examination of the validity tests themselves. Are some validity tests either logically or empirically more important than others? Is there a tradeoff between some of these tests, for example, extending the work of Campbell (Campbell, 1966) that improving internal validity comes at the expense of decreasing external validity?

Discussion of validity tests is essentially a discussion of the epistemological question – "how do I know that know?" Or, to make a less strong claim, "how do I know that the knowledge generated by this model is something that I can have some confidence in?"

But there are practical benefits in the ontological question too. The ontological question asks "what is the nature of the reality of the things that I am interested in gaining knowledge of"?

Various views of this could be that an "organization" or the "public interest" is real as a social construct, and that it has no independent existence. Others may argue that organizations or the public interest must have an existence independent of what anyone individual or group believe it is. Once you have identified something's ontological status, you can begin to answer the epistemological question, "how do we know something about that thing? We need to know the nature of the reality of things (how does it exist) before we can know what knowledge we can have of it.

The benefit of understanding this distinction, is that practically, most of the things we work with in public life could be described in either of these ways. We do not need to prove whether the "public interest" is a social construct in all situations, only that if in a local situation, it looks like there is relatively less agreement versus more agreement on what we as a society should do, that for all practical purposes, in that local situation, it will have to be something that is constructed. And, as a result, the logic of the social construction description of reality would be important to understand and employ.

So, simulation modelers do not need to answer these tough philosophical questions. But they need to know the questions because those questions are not merely questions made up by philosophers (Schmid, 2005). They are questions that anyone thinking about how to "produce" good knowledge probably has pondered to some degree. Applying this to simulation, what is the nature of the "rules" which produce complex phenomena that are emergent, local, time-dependent, and non-linear? And derivatively, what kind of knowledge can we say exists in that simulation? For example, some may argue that causation is implicit in systems dynamics models (Barlas, 1994), but not all modelers would hesitate to assign that much confidence to their models. Yucel (2009) identifies a number of different policy models that can be assisted through simulation. Some of them are clearly of the sort that involves representing a reality, while others, could merely explore the logical implications of the assumptions made by decision-makers, clarifying the assumptions of a group, and maybe even negotiating solutions. To what extent are each of these activities modeling "real" things? How are they real? And more importantly, how do we know that we have obtained good knowledge about them? In conclusion, exploring how people view what simulation is, and what its contribution is, can go far in helping with the very practical problem of making it more useful. We move from questions about ontology to the epistemological question about how much confidence we should vest in a model, namely, how valid is that model?

Simulation and Social Systems

To say that a model is validated is to say that the model is, within sound bounds of uncertainty, an accurate representation of the real world (Logan and Nitta, 2002). In policy analysis, the likelihood of differentiating the independent effect of some input on some outcome is quite low. Social science, in general, does not lend itself well to full explication of the complexities of any system in which we have interest. When systems are sufficiently closed (or closable), validating models is a relatively straight-forward process of controlling settings and replicating models to determine if the same results are achieved. Policy systems, however, are never closed (Kleindorfer, et al, 1998), and the complexity inherent in open systems brings verification and validation issues to the forefront (Fischer, 2006). A policy model must pass a host of validation hurdles, comparing the model with the system of which it is a representation, before it is useful, either for purposes of creating knowledge or informing policy decision making (Worren, et al, 2002). In this document we discuss issues of validation in policy modeling, describe the concepts that are addressed in the validation process and propose a comparative and chronological procedure for increasing the validity and usefulness of models of public policy issues.

Validation Tests in Simulation, Psychometrics, and Statistics

There are multiple validation issues requiring attention at each step of the process of creating a policy model. Models go through three stages during the development process: pre-development, development, and model-use (Yucel and van Daalen, 2009). Planning occurs during the pre-development phase, where the chief validation concerns relate to the logical flow of the system being modeled, the model's theoretical basis and how to simplify the complexities of the system in order to create a workable model

(Kleindorfer, et al, 1998). At a practical level, the concern during this phase relates to the model's structure. During the development phase, concerns center on the extent to which the model reflects perceived reality and the model's sensitivity to circumstantial changes (Sargent, 2000). The practical consideration at this stage relates to the model's behavior. In the model-use phase, considerations turn to practical concerns regarding the implications and usefulness of the model (Worren, et al, 2002). Figure 1 provides a schematic view of the model development phases.

There are multiple ways to think and talk about validity and the process of validation. There appear to be at least two communities that do not always talk and therefore use different concepts and jargon; the psychometric/statistical view of validity (e.g., Hair, et al, 1998) and the systems/simulation view (e.g., Forrester and Senge, 1980). (See also McNamara, Trucano, Backus, Mitchell, & Slepoy, 2008; Scheiber, 2002). In general, however, these differences are superficial; in fact, we believe that there is a general correspondence in the concepts and the jargon. However, there are also are differences, and these appear to be opportunities for useful exchange. We therefore incorporate validity concepts from both traditions as we describe the validation process through each of the three stages of model development. At each stage in this process, we can construct hypothesis tests to clarify the nature of the validity that is being attained (Sargent, 2000).

Pre-development

During the pre-development phase, the primary focus is on the extent to which a proposed model will, as a simplification of reality, accurately reflect a real-world system. Yucel and van Daalen (2009) describe validity at this stage as including the setting of the boundaries of the model. Policies function in open systems, but in order to study and learn about possible policy outcomes, modelers must choose some level at which to artificially close the system. With this constraint in mind, the first question is how to bound the system, and in that context, identify the point at which a model will be considered "valid".

As an aside, while some imply that validity is a dichotomous yes or no (Oreskes, et al, 1994), most consider validation to be scaled (Gass, 1983; Balderstone, 1999; Forrester and Senge, 1980). That is, there is never a point at which an "invalid" model switches to being "valid." In many respects this question really depends, again, on the user. For decision-makers, a model could be considered valid and will be used, or dismissed. Modelers, on the other hand, or continually work with refining a model, are probably more comfortable with the notion of a scaled view of validation.

Either way, as models pass validation tests that increases user confidence, models achieve "greater" validity the more they are seen as accurately reflecting real systems (Barlas, 1996). As part of the pre-development phase, criteria must be established, together with the level of confidence for certifying a model to be "valid enough."

Relevance is the overriding concern for addressing policy questions (Fischer, 2006). If a model of a policy issue is not relevant for the problem it is meant to address, then it immediately lacks pragmatic validity (Worren, et al, 2002). Models that inform policy decisions are only as useful as they are relevant and implementable. A defining feature of policy analysis is this concern with the *pragmatic relevance* test; interesting or even theoretically/scientifically valid models may be of little use in a policy context if their implementation is infeasible or they are perceived to be of little use to policymakers (Fischer, 2006). *Pragmatic* validity should be considered early and often throughout the modeling process, and it is a theme to which we will return.

Other specific validity issues at the pre-development stage center on the theoretical basis for the model, and whether the structure of the model corresponds to the structure of the actual system. To begin, modelers should consider the *face validity* of the proposed model. A model must be a plausible abstraction for it to have face validity; to the extent that propositions regarding behavior in the system are plausible, the model has face validity. If a model substantially contradicts what we collectively agree is possible, it will be viewed as lacking face validity.

We consider the *construct validity* of a model when we assess the credibility of the link between model assumptions and theory. In any model, assumptions should be based on theory. While in a traditional statistical model these assumptions are often in the background, in a simulation model, assumptions come to the forefront. In simulations, assumptions regarding the underlying constructs and the interactions among these constructs must be explicitly modeled and stated, enabling model users to have a clear view of construct validity. If the link between model assumptions and theory is clear and consistent, then the model likely will be viewed as having sufficient construct validity. If model assumptions are *ad hoc* or based solely upon modeler expectations absent some grounding in received wisdom and research, the model is likely to be of limited usefulness, as it is unlikely to provide any insight into broader system behavior.

Face and construct validity are important whatever model one is using, but *context validity* is an extremely important, yet often overlooked aspect of policy models (Fischer, 2006). Achieving sufficient face validity or construct validity is not difficult. Having a plausible model that is based on some previous knowledge is an important, but low bar to attain. Determining the context of a model is more of a challenge, partly because it depends of the objectives of and the motivation for the modeling activity. There is usually much less clarity about what is the relevant context and is therefore difficult to ascertain whether the bar has been met. When defining the context of a model, we are setting the boundaries of the system being modeled, determining what relationships are to be part of the model and which are not (Yucel and van Daalen, 2009). When considering public policies, context is particularly important. Local conditions and considerations are widely variable, and policy outcomes in one context can differ considerably from outcomes in another context. As previously stated, a policy model is only useful if it is relevant to the policy problem being considered (Fischer, 2006). Therefore, the context for the model must be relevant to the problem, and this context is a negotiated outcome that can be found via a thorough understanding of the problem, analysis of previous research in the area and through consultation with stakeholders, experts and policy makers (Richardson and Anderson, 1994).

Along similar lines, in the pre-development phase, in addition to considering which relationships to include, careful attention must also be paid to the parameters and variables included. When determining the context, we consider which variables and parameters to include or exclude; a model has *content validity* to the extent that the variables and parameters chosen to be included are accurate reflections of those variables and parameters in the real system. Thus, as we begin to develop the actual model, attention turns to the data upon which a model will be based. Whether developing an empirical model based on observed data, or a simulation, those data must be valid representations of the actual characteristics they are meant to reflect.

In the pre-development phase, validity concerns relate to the decisions modelers make as they develop simple representations of complex real systems. The predominant concerns are whether the model will accurately reflect the underlying system and provide information that is seen to be useful both in terms of theory and practice. As the model begins to take shape, focus shifts to more technical validation issues, but the pre-development considerations carry throughout the modeling process, and in fact, it is an iterative process.

Development

As a model is being developed, validity concerns shift focus more directly to the model itself. The work of Forrester and Senge (1980) and Barlas (1996) provide the basic framework for checking model validity during the development stage. Using the psychometric nomenclature, most of the proposed validity tests, at this stage, are of *convergent* (or *concurrent*) validity. The model should converge on instances where the model predicts behavior that has been observed. Confidence is built as the model increasingly shows that it is able to accurately retrodict what has actually happened in the real system. By comparing model results with past, observed behavior in the actual system, the model can gain credibility over time. Although the ability to accurately re-create the past is no guarantee of being able to predict the future, accurate retrodiction is seen as an indicator of model validity.

Although no model is ever completely validated, the tests during this phase increase the level of confidence users have in the model results. If the model adequately meets the validity criteria, then there is likely to be more confidence, which makes the model more useful. Forrester and Senge (1980) detail a battery of tests for building confidence in a model. The tests cover both verification and validation and extend through all three phases of model development, but are primarily focused on post hoc procedures. While building the model itself, the procedure should adhere to the decisions made in the pre-development phase, verifying the model as it is being developed.

The Forrester and Senge (1980) tests fall into three broad areas: model structure, model behavior, and policy implications. In essence, by validating that the structure of the model matches the structure of the real system and the behavior predicted by the model matches behavior in the real system, we ensure that the model will be viewed as a relatively valid representation of the real system. To that end, the Forrester and Senge (1980) tests are comparative in nature, with the expectation that a model is more valid to the extent that it compares favorably to the structure and behavior of the real system. The Forrester and Senge (1980) tests are briefly described in Table 1, according to the category of validity being considered.

--- Insert Table 1 about here ---

The *structure verification* test is essentially a test of *convergent* validity, whereby the structure of the model is compared with the structure of the real system, and the parameter verification test is akin to content validity. Assuming that boundary adequacy was addressed prior to development, and that unit conformance has been verified (that is, the results are not presented in meaningless units like people squared), the key insight for model structure is the *extreme conditions test*. Subjecting the model to extreme values of the parameters checks the robustness of the model's predictions. If the model fails to function at extremes, then its structure is not adequate.

The crux of validation in the model development phase is in the model behavior area, where tests are centered on ensuring that the behavior in the model adequately conforms to the behavior of the real world system being modeled. A model gains confidence for each test it passes; the more it is able to predict actual system behavior, even under extremes or unexpected changes, the more confidence users can have that the results are a reasonable facsimile of the behavior of a system given certain policy changes. We will return to the policy implication of tests in the next section.

Forrester and Senge (1980) present these tests conceptually, while Barlas (1996) attempts to provide a systematic procedure for achieving a level of confidence and validity. Barlas' (1996) recommendations fit within the Forrester and Senge (1980) framework, suggesting a series of statistical tests comparing model predictions with real system observations. The closer that model predictions correspond with real world observations, the more confidence we can have in the model's predictions for future unknown behavior.

Broadly, at the development stage of model building, concerns with validity relate to the model's behavioral predictions and sensitivity. The main consideration is ensuring that, as the model is being built, it is being calibrated to the behavior of the actual system that is being modeled. The key under-addressed issue at this stage is that past performance does not necessarily guarantee future prediction. By calibrating a model with observed data from the past, we ensure that the model is a valid representation of past system behavior, but one of the key complications for modeling social science behaviors is that past behavior is an uncertain predictor of future behavior. It is a probabilistic predictor to be sure, but it is far from certain that modeling assumptions that accurately predict past behavior will lead to accurate predictions of future behavior.

Model use

The Forrester and Senge (1980) tests of policy implications provide a good starting point for considering validation during the model use phase, where concerns turn to the implications and interpretations of the model's results. Although Forrester and Senge (1980) never define what they mean by "system improvement" (improvement for whom? how?), it is conceptually useful to consider whether the model is able to help us identify those policy changes that will be an improvement, however defined. Model predictions of future behavior always have a level of uncertainty. The *policy sensitivity test* acknowledges this uncertainty; the level of uncertainty is particularly important for the *pragmatic validity* of policy model results. If a model sufficiently satisfies all the previous validity criteria, but the results are highly uncertain, then it is likely to be of limited usefulness to policy makers.

Taking a step back, to be useful for policy makers, model results must have some *implications for policy*. A perfectly valid model indicating that the variables that are important to system outcomes are not those that can be affected by policy (say, ethnic composition of a given community, for instance) is of no use to policy makers. Further, if important variables can only be changed in ways that are impractical in the real system, the model is, at best, an interesting experiment.

Modelers must also consider *consequential* validity. A model that provides useful information, that informs both theory as well as policy makers, must also indicate that the consequences are valid and socially acceptable. There are several important questions modelers must ask regarding the consequences of simulation findings. How are test results likely to be used? Will the consequences of policy changes in the system of interest potentially lead to unacceptable unintended consequences in other systems? Could the results be used in an unethical manner?

Finally, modelers should consider the *scientific validity* of the model results. Ideally, models should contribute to knowledge and theoretical understanding of the system in question. If a model is applicable to only a few specific scenarios, then the results are not generalizable and do not provide us with a deeper understanding of the system in question. Table 2 is a schematic of the model development phases linking the Forrester and Senge (1980) test categories with the phases of the development process and different types of validity that are associated with each stage and category.

--- Insert Table 2 here ---

Practical considerations

In general, determining model validity is a comparative exercise. We compare models with similar research, with real world systems and with our expectations. Therefore, it is important to consider how to establish the criteria against which a model will be compared. Procedurally, most experienced modelers recommend that model verification should be carried out by a technically proficient third party (i.e., testers who are technically proficient with the modeling process/software being used). The comparison for verification is ensuring whether results calculate according to the applicable mathematical rules.

The comparison for validation is more complex to attain. By and large the comparison is with some aspect of behavior in the real system. This comparison can be assessed by combing though relevant research and establishing a full set of expectations regarding system behavior. A common approach in system dynamics modeling is to consult and include experts in the model development process (Richardson and Anderson, 1994). Through a participatory, deliberative process, subject matter experts can arrive at agreed upon expectations for the behavior of the actual system that can inform the behavior of the model.

Conclusion

Ensuring policy model validity is a process of comparing simplified models with real world systems. However derived, actual system behavior can only be estimated. As such, validity can never be absolute, with a distinct tipping point at which an invalid model becomes valid. Model validity is an ongoing process of building confidence that a model is an accurate reflection of the real world system it is meant to represent. In this document, we presented a systematic conceptual process for assessing the validity of a policy model. Validity testing takes place in three phases of the model development process: pre-development, during development and during model use. At each stage, model assumptions and behaviors are compared with those of the real world system, presented here in the context of traditional views of validity from psychometrics combined with system dynamics procedures primarily associated with Forrester and Senge (1980). While a model that sufficiently passes the conceptual hurdles described here is not necessarily "valid", it is likely that users will view model results more confidently and will more fully utilize results than otherwise would be the case.

Table 1: Forrester and Senge's (1980) Tests for Building Confidence in Models

Category	Test	Description
Model Structure	Structure-verification	Does model structure correspond with structure of real system?
	Parameter-verification	Do model parameters correspond conceptually and numerically with real system parameters?
	Extreme-conditions	Does the model remain stable under extreme variation?
	Dimensional-consistency	Do model units of measure correspond with actual units?
	Boundary-adequacy	Does the structure satisfy the model's purpose?
Model Behavior	Behavior-reproduction	Does behavior in the model match observed behavior in the real system?
	Behavior-prediction	Does the model predict plausible behavior in the future?
	Behavior-anomaly	Do model assumptions produce any behaviors that are anomalous to behavior in the real system?
	Family-member	Is the model generalizable to related systems or is it only applicable to one unique situation?
	Surprise-behavior	Does the model predict the same unexpected behavior that has occurred in the real system?
	Extreme-policy	If an extreme policy is followed in the real system, does the model accurately predict the actual outcomes?
	Boundary-adequacy	Are all necessary elements in place for the model to address the policy problem?
	Behavior-sensitivity	Do plausible changes in model parameters lead to implausible results?
Policy Implications	System-improvement	Does the model provide results that are beneficial to policy makers?
	Changed-behavior-prediction	Does the model accurately predict changes in behavior brought about by changes in policy?
	Boundary-adequacy	If model structure is changed, do the policy recommendations change dramatically?
	Policy-sensitivity	How uncertain are the policy recommendations?

Table 2: Schematic of model development phases

Development Stage	Pre-Development	Development	Model Use
Validity	*Face* *Context* *Construct*	*Concurrent* *Content* *Procedural*	*Predictive* *Consequential* *Pragmatic* *Scientific*
Test Category	STRUCTURE	BEHAVIOR	IMPLICATIONS

References

Balderstone, S. 1999. Increasing User Confidence in System Dynamics Models Through Use of an Established Set of Logic Rules to Enhance Forrester and Senge's Validation Tests. *Proceedings of the 17th International Systems Dynamics Conference.*

Barlas, Y. 1994. Model Validation in System Dynamics.*1994 International System Dynamics Conference*.

Barlas, Y. 1996. Formal Aspects of Model Validity and Validation in System Dynamics. *System Dynamics Review*. 12(3): 183-210.

Campbell, D. T., -. (1966). *Experimental and quasi-experimental designs for research [by] Donald T. Campbell [and] Julian C. Stanley*. Chicago: R. McNally.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes, 108*(1), 116-130.

Dunn, W. 2002. Using the Method of Context Validation to Mitigate Type III Errors in Environmental Policy Analysis. In Hisschemoller, M., Hoppe, R., Dunn, W., and Ravetz, J. (eds) *Knowledge, Power and Participation in Environmental Policy Analysis*. Edison, NJ: Transaction Publishers.

Faust, David. (1982). A Needed Component in Prescriptions for Science: Empirical Knowledge of Human Cognitive Limitations. Knowledge 3(4): 555-570.

Fischer, F. 2006. *Evaluating Public Policy*. Thompson: Tampa, FL.

Forrester, J., and Senge, P. 1980. Tests for Building Confidence in System Dynamics Models. *TIMS Studies in Management Sciences*.14: 209-228.

Gass, S. 1983. Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Operations Research*. 31(4): 603-631.

Kleindorfer, G., O'Neill, L., and Ganeshan, R. 1998. Validation in Simulation: Various Positions in the Philosophy of Science. *Management Science*. 44(8): 1087-1099.

Logan, R. and Nitta, C. 2002. Verification & Validation (V&V) Methodology and Quantitative Reliability at Confidence (QRC): Basis for an Investment Strategy. *Report prepared for U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory*.

McNamara, L. A., Trucano, T. G., Backus, G. A., Mitchell, S. A., & Slepoy, A. (2008). R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation. In Energy (Ed.) (Final Report ed.). Albuquerque, NM: Sandia National Laboratories.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. 1994. Verification, Validation and Confirmation of Numerical Models in the Earth Sciences. *Science*. 263: 641-646.

Repenning, N. P. (2003). Selling system dynamics to (other) social scientists. *System Dynamics Review, 19*(4), 303-327.

Richardson, G. and Anderson, D. 1994. Teamwork in Group Model Building. *System Dynamics Review*. 11(2): 113-137.

Sargent, R. 1999. Validation and Verification of Simulation Models. *Proceedings of the 1999 Winter Simulation Conference*.

Sargent, R. 2000. Verification, Validation, and Accreditation of Simulation Models. *Proceedings of the 2000 Winter Simulation Conference*.

Scheiber, D. (2002). Validating Agent-Based Models: From Metaphysics to Applications, *Midwestern Political Science Association* (pp. 19). Chicago.

Schmid, A. (2005). What is the truth of simulation? *Jasss-the Journal of Artificial Societies and Social Simulation, 8*(4), 7.

Sterman, J. D. (1987). Testing Behavioral Simulation Models by Direct Experiment. *Management Science, 33*(12), 1572-1592.

Worren, N., Moore, K., and Elliott, R. 2002. When Theories Become Tools: Toward a Framework for Pragmatic Validity. *Human Relations*. 55(10): 1227-1250.

Yucel, G. and van Daalen, E. 2009. An Objective-Based Perspective on Assessment of Model-Supported Policy Processes. *Journal of Artificial Societies and Social Simulation*. 12(4): 3-29.