Developments in Business Simulation and Experiential Learning, Volume 28, 2001

FIDELITY, VERIFIABILITY, AND VALIDITY OF SIMULATION: CONSTRUCTS FOR EVALUATION

Andrew Hale Feinstein, UNLV College of Hotel Administration
andyf@nevada.edu

Hugh M. Cannon, Wayne State University
HughCannon@email.com

ABSTRACT

Although instructional research on simulation has been around for almost 40 years, very little headway has been made on the creation of an academically acceptable methodology for evaluating this type of learning environment. Several comprehensive reviews of simulation assessment literature have all concluded that this problem stems from poorly designed studies, the lack of a generally accepted research taxonomy, and no well defined constructs with which to assess learning outcomes. In an effort to assist in the development of a simulation evaluation taxonomy, this paper focuses on identifying, defining, and explaining key concepts within three of the most important constructs of simulation evaluation – fidelity, verifiability, and validity. *KEY WORDS: simulation fidelity, simulation verifiability, simulation validity*

INTRODUCTION

One of the major problems of simulations is how to “evaluate the training effectiveness [of a simulation]” (Hays and Singer, 1989, p. 193). Although for more than 40 years, researchers have lauded the benefits of simulation (Wolfe and Crookall, 1998), very few of these claims are supported with substantial research (Butler et al., 1988; Miles et al., 1986; Wolfe, 1981, 1985, 1990; Wolfe and Crookall, 1998). Part of the problem is that "effectiveness" depends on a simulation's purpose. However, much more research has gone into simulation development than evaluation. Many of the researchers cited above attribute this lop-sided effort to poorly designed studies and difficulties inherent in creating a methodology of evaluation. The irony of this problem is that taxonomies of evaluation must be developed if we are to understand what an effective simulation is, regardless of its purpose. Further, constructs must be clearly defined if we are to determine a simulation's appropriate use and create the methodology to measure its effectiveness as a learning environment. This importance is clearly explained by Wolfe and Crookall (1998, p. 8):

> The educational simulation/gaming field has been unable to create a generally accepted typology, let alone taxonomy, of the nature of simulation/gaming. This is unfortunate because the basis of any science is its ability to discriminate and classify phenomena within its purview, based on underlying theory and precepts.

Research in simulation has focused primarily on fidelity, verification, and validation as constructs of effectiveness assessment. Although researchers are mostly in agreement regarding the processes of verifying the operational aspects of a simulation model, they have yet to agree on what defines an effective measurement device for validating a simulation or what is an appropriate level of fidelity for this type of learning environment.

The purpose of this paper is to review the current literature on simulation fidelity, verifiability, and validity – three of the most important constructs of simulation evaluation – and to identify, define, and explain key concepts within these domains. It is hoped that this effort might guide future research efforts focused on simulation evaluation.

SIMULATION FIDELITY

Fidelity is the level of realism that a simulation presents to the learner. This concept is an integral component in simulation because it defines “how similar a training situation must be, relative to the operational situation, in order to train most efficiently” (Hays and Singer, 1989, p. 1).

The term fidelity “has most often referred to the design of simulators that are used in training ” (Hays and Singer, 1989, p. 47). Further, Hays and Singer point out that it “should be restricted to descriptions of the required configuration of the training situation and not to be used when discussing behaviors” (1989, p. 47). Fidelity focuses on the equipment that is used to simulate a particular learning environment. These authors (1989, p. 50) sum up these concepts by defining fidelity as:

Best Paper Award Recipient 2001 - Simulation Track

> …the degree of similarity between the training situation and the operational situation which is simulated. It is a two dimensional measurement of this similarity in terms of: (1) the physical characteristics, for example visual, spatial, kinesthetic, etc.; and (2) the functional characteristics, for example the informational, stimulus, and response options of the training situation.

FIDELITY IN TRAINING AND EDUCATION

The degree of fidelity in a learning environment is an extremely difficult element to measure. Many authors studied the relationship between fidelity and its effects on training and education in the 1960s and 70s (Greenlaw, Herron, and Rawdon, 1962; Kibbee, 1961; Miller, 1978; Martin and Waag, 1978; Cox, Wood, Boren, and Thorne, 1965; Muckler, Nygaard, O'Kelly, and Williams, 1959; Blaiwes and Regan, 1986; Bunker, 1978; Kinkade and Wheaton, 1972). These studies found that a higher level of fidelity does not translate into more effective training or enhanced learning. In fact, many studies found that lowered fidelity actually can assist in acquiring the details of training and education (Gagne, 1954; Dwyer, 1974; Miller, 1974; Alessi, 1988). Further, Martin and Waag (1978) determined that high fidelity can actually hinder effective training and learning because it overstimulates novice trainees.

Blaiwes and Regan (1986) believe that in simulation the goal is to provide a learning environment, not a vehicle for trainees to exhibit perfect performance. Kibbee (1961) believes that most models are designed to show general principles and that a player's perception of verisimilitude is far more important than the similarity of a model to the real world. Although many of these articles may seem to be outdated, the fidelity theories and principles that emerge from them are the foundation that current simulation modelers rely upon for creating effective learning environments.

Hays and Singer (1989) point out that it can be cost effective for novice trainees to utilize low fidelity devices during the first stages of learning. These authors also believe that a simulator does not need to be an exact representation of the real world in order to provide effective training. In fact, they feel that it may be necessary to "depart from realism in order to provide the most effective training" (Hays and Singer, 1989, p. 15). They also believe that some components of the simulator – such as being able to stop or restart the model, or a refined feedback mechanism – would "reduce the realism of the training situation, but enhance learning." (Hays and Singer, 1989, p. 15).

SIMULATION VERIFICATION

The general notion of validation incorporates two different processes: *verification* and *validation* (Pegden, Shannon, and Sadowski, 1995). *Verification* is the process of assessing that a model is operating as intended. *Validation* is the process of assessing that the conclusions reached from a simulation are similar to those reached in the real-world system being modeled. In other words, "Validation is the process of determining that we have built the right model, whereas verification is designed to see if we have built the model right" (Pegden, Shannon, and Sadowski, 1995, p. 129).

The process of verification involves debugging the model by isolating and eliminating as many errors as possible. This can be done by using internal debuggers of the simulation software, viewing output reports, evaluating step-by-step traces of a simulation run, and involving individuals who can evaluate the simulation. Many verification errors are simple problems of software debugging. Others involve fixing design errors, where the basic equations interact in unanticipated ways, or where the embedded response functions become invalid for extreme values.

Needless to say, it is very important to remove all errors in a simulation to ensure it is operating as intended. The isolation of errors in a simulation model can be an extremely difficult task. Therefore, it is vital to use various methods for identifying and eliminating errors. Typically, this testing is done first with alpha tests by the simulation developer and later, with beta tests, where the simulation is run in a variety of conditions by independent users.

SIMULATION VALIDATION

THE PROBLEM OF VALIDATION: AN OVERVIEW

Over the past several years, the Association for Business Simulations and Experiential Learning (ABSEL) Assessment Committee has been engaged in a project aimed at evaluating and registering simulation games, as a means of supporting teachers and consultants in their efforts to find simulations that work properly. This initiative has raised more issues than it has resolved. But raising issues is precisely what was needed. Figure 1 provides an overall picture of the problems the Committee faced.

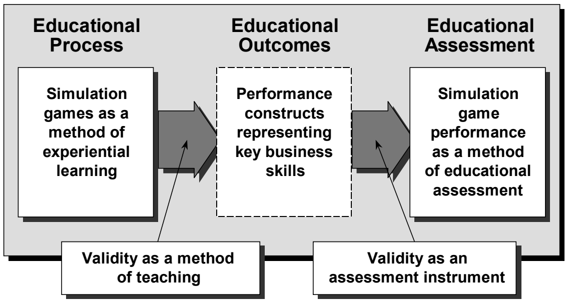

Figure 1. Two Faces of Simulation Game Validation From Phil Anderson, Hugh M. Cannon, Dolly Malik and Precha Thavikulwat (1998), "Games as Instruments of Assessment," in John K. Butler, Jr., Nancy H. Leonard and Sandra W. Morgan (eds.), Developments in Business Simulation and Experiential Learning, V. 25, pp. 31-9.

Regardless of the purpose for which a simulation is being validated, at one level, the basic issues are very similar. Table 1 summarizes some of the key concepts used in the literature related to business simulation. These provide a basic vocabulary for understanding simulation validation.

Table 1
Concepts Related to Simulation Validation Research

Concept	Definition	Reference
Accuracy	Does a simulation game accurately mirror the reality it is supposed to represent. (A type of *external representational validity*. See *event empirical validity*, *event validity*, and *realism*).	Dukes and Waller, 1976
Algorithmic validity	Does the model return appropriate values? (A type of *internal representational validity*).	Wolfe and Jackson, 1989
Believability	Does the simulation model's ultimate user have confidence in the model's results? (Reflection of perceived *internal* or *external representational validity.* A key issue in establishing *internal educational validity*).	Pegden, Shannon, and Sadowski, 1995
Conceptual (face) validity	Does the model adequately represent the real-world system? (Special case of *external representational validity.* A key issue in establishing *internal educational validity*).	Pegden, Shannon, and Sadowski, 1995
Construct validity	How correctly are the variables in the model related to each other to form strategic and environmental constructs? (A special case of *internal representational validity*)	Carmines and Zeller, 1979; Babbie, 1992, pp. 132-133
Content validity	How complete is the simulation model? (A special case of *internal representational validity*).	Carmines and Zeller, 1979; Babbie, 1992, pp. 132-133
Criterion (predictive) validity	Does the model effectively predict real-world situations? (Special case of *external representational validity*. See *accuracy*, *realism*, *empirical validity*)	Carmines and Zeller, 1979; Babbie, 1992
Educational validity	Does the simulation provide a valid learning experience and/or assessment of learning? (As contrasted with *representational validity*)	Conceptualized in this paper

Empirical validity	Does a simulation game exhibit a closeness of fit to other measures of the phenomena it is designed to simulate. (A type of *representational external validity*. See *event accuracy*, *event validity*, and *realism*).	Boocock, 1972
Empiricism (objectivism, foundationalism, verificationalism)	The philosophy that the determination of validity must ultimately be determined by the analysis of objective data. (A naive philosophy for externally validating simulation).	Weinberg, 1936
Event validity	The degree to which a simulation's predicted responses correspond to actual data from the organization being simulated. (A type of *external representational validity*)	Mihram, 1972
External validity	Does the simulation model represent actual external phenomena? (Applicable to issues of both *representational* and *educational validity*).	Cook and Campbell, 1979
Hermeneuticism	The belief that meaning does not sit in the object or its interpreter; rather it is generated through an interaction that is guided by phronesis, or an inherent sense of what is right. (A guiding philosophy for establishing *internal representational*, *educational, and external validity*)	Bernstein, 1983; Gadamer, 1976
Internal validity	Do a model's relationships represent true causality? (An issues relating to *external representational validity*).	Cook and Campbell, 1979
Operational validity	Are the model-generated behavioral data characteristic of the real-world system's behavioral data? (Special case of *external representational validity*)	Pegden, Shannon, and Sadowski, 1995
Plausibility	Does the simulation model appear to represent real-life phenomena? (Reflection of perceived *external representational validity.* See: *accuracy*, *believability*, *criterion* or *predictive validity*, *plausibility*, *verisimilitude*).	Boocock, 1972
Positive Economics	An approach to validation where scientific laws are only organizers of experience and used to make propositions about empirical observations. (A naive philosophy for externally validating simulation).	Friedman, 1953
Rationalism	The belief that we all have access to some fundamental set of principles that are part of our cognitive apparatus. (A naive philosophy for externally validating simulation).	Decartes,1993; Kant, 1996
Realism	Does the simulation represent the business environment it is designed to simulate (A type of *external representational validity*. See *event validity*, *accuracy*, and *empirical validity*).	Norris, 1986
Relativism (Conventionalism)	The belief that one may not have access to any foundational principles – one's opinion is most important. (A naive philosophy for externally validating simulation).	Feyerabend, 1993; Kuhn 1962; Popper, 1959

Representational validity	Does the simulation provide a valid representation of a desired phenomenon?	Conceptualized in this paper
Validation	As opposed to *verification*, *validation* asks whether the model is correct.	Pegden, Shannon, and Sadowski, 1995
Verification	Does the model do what it intends to do.	Pegden, Shannon, and Sadowski, 1995
Verisimilitude	Does the simulation model appear to represent real-life phenomena? (Reflection of perceived *external representational validity* See *accuracy*, *criterion* or *predictive validity*, *plausibility*, *believability*).	Kibbee, 1961

INTERNAL VALIDITY

Validation is typically more demanding than verification. Campbell and Stanley (1963) divide the general concept into two types of validity: *internal* and *external*, as suggested by Figure 2. Cook and Campbell describe these types of validity:

> Internal validity refers to the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause. External validity refers to the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times. (Cook and Campbell, 1979, p. 37).

Figure 2 suggests two forms of internal validity in a simulation game. The first relates to the logic and structure of the game itself – following what we will refer to as *representational validity*. It asks the question, 'to what extent does a simulation game accurately represent desired phenomena?' For instance, in a marketing simulation, do advertising expenditures actually contribute to demand in some reasonable manner? Do they interact with the nature of the product, rewarding consistency between the appeal and actual consumer benefits delivered? Do strategically related decisions hang together in a recognizable manner? And so forth. These address what Wolfe and Jackson (1989) call *algorithmic validity*. That is, does the algorithm used in the simulation really model the phenomena it is supposed to represent. Does the algorithm return values that follow the intended pattern? In a similar vein, one might ask how complete the model is, or what some theorists have referred to as *content validity*, and how correctly the variables are related to each other to form strategic and environmental constructs i.e., *construct validity* (Carmines and Zeller, 1979; Babbie, 1992, pp. 132-133)?

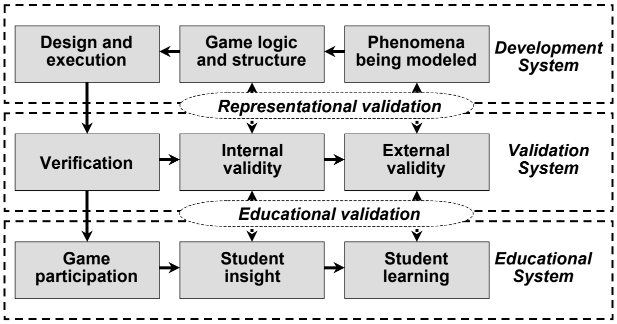

Figure 2. Two Patterns of Simulation Game Validation

The second form of internal validity addresses the degree to which game participants understand the game and play it with insight – following what we will refer to as *educational validity*. That is, to what extent are student decisions influenced in the intended manner by game design (Parasuraman, 1981). The logic of an educational simulation is that it will provide a learning environment in which students can observe modeled phenomena and develop managerial insight to address them. In order to achieve internal educational validity, game participants would have to discern the phenomena being modeled. This test is analogous to a "manipulation check" in experimental research.

Consistent with this approach, Dickson, Whitely and Faria (1990) addressed internal validity as the degree to which students tend to recognize and then select a promotional strategy (e.g., "push" versus "pull") appropriate to the simulated environment. If students did not recognize that the game rewarded one pattern of promotion more than another, we would conclude that decisions were being made randomly and that the game was not internally valid from an educational standpoint.

Note that many researchers have equated internal validity with the educational effectiveness of a simulation (Bredemeier and Greenblat, 1981; Greenlaw and Wyman, 1973; Norris, 1986; Pierfy, 1977; Wolfe, 1985). This is true for educational validity, in the sense that it implies a student understanding of the phenomena being simulated. However, it is not necessarily true for representational validity. Indeed, a simulation that faithfully represents strategic cause and effect might well be so complex that students never see the relationships (Cannon, 1995). Furthermore, the relevance of any understanding students have of the simulation to the real world of business is a question of external validity. A simulation that taught students patterns and relationships that were out of sync with the real world of business could be internally valid, but externally disastrous.

EXTERNAL VALIDITY

Figure 2 suggests that the logic and structure of the game are reflections of some real world phenomenon. This is a reflection of external validity (Boocock, 1972; Dukes and Waller, 1976; Mihram, 1972; Norris and Snyder, 1982; Wolfe and Roberts, 1986, 1993). Simulation models are typically designed as representations of pre-existing systems or systems that are under consideration. In order to be able to use simulation as a tool to assess characteristics of a real-world situation, the simulation model must be an effective replica of that system. This replica needs to contain only a "degree of homomorphism" (Stanislaw, 1986, p. 177; Vandierendock, 1975) between itself and the system it is modeling, commensurate with a set of objectives. The ability for developers of simulation to prove that their models replicate the real world at some level is termed the problem of simulation validation.

Whereas most researchers tend to agree on the nature of internal validity, there is considerable disagreement regarding external validity. However, much of the problem can be attributed to the fact that the validity of the simulation is keyed to its objectives. This is represented in Figure 2 by the distinction between representational and educational validation. For instance, Mehrez, Reichel, and Olami (1987) and House and Napier (1988) study the degree to which simulated companies behave like real ones. Related to this is Boocock's (1972) notion of *empirical validity*, Mihram's (1972) *event validity*, Dukes and Waller's (1976) *accuracy*, Norris' (1986) *realism*, and *criterion* (*predictive*) *validity* (Carmines and Zeller, 1979; Babbie, 1992). Each of these is concerned with the degree to which a simulation behaves in ways that are similar to the organizations and markets they represent.

All of these terms refer to representational validation. While their conclusions are useful in studying educational processes, they are nevertheless quite different from studies of educational validation. As we see in Figure 1, the desired output from an educational simulation is not an accurate replication of what would happen in the real world at all, but rather a set of skills that will help students make real-world decisions. In this context, then, external validation means either the demonstration that a simulation teaches key business skills (validation as a method of teaching), or that key business skills are needed in order to perform well in a business simulation game (validation as an assessment instrument).

In place of concerns for representational validity, theorists speak of *verisimilitude* (Kibbee, 1961), *plausibility* (Boocock, 1972), and *believability* (Pegden, Shannon, and Sadowski, 1995). These terms do not represent any form of validity, but only the perception of it. The most direct implication is on student motivation and insight – both of which impact most directly on internal validity.

EXTERNAL VALIDITY AND THE UNDERLYING PHILOSOPHY OF SCIENCE

> The [external] validation problem in simulation is an explicit recognition that simulation models are like miniature scientific theories. Each of them is a set of propositions about how a particular manufacturing or service system works. As such, the warrant we give for these models can be discussed in the same terms that we use in scientific theorizing in general. (Kleindorfer, O'Neil, and Ganeshan, 1998, p. 1087)

The problem of external simulation validation has been prevalent since the inception of simulation and raises many philosophical questions. Although there is a "vast amount of literature about the concept of validity…this literature focuses mainly on the validity of experimental situations or on the validity of measurement instruments. The concept of validity in relation to simulation as a simplified model of a complex reference system is hardly elaborated in the literature" (Peters, Vissers, and Heijne, 1998, p. 23).

So how can simulation modelers 'prove' that their models replicate real-world systems? What methodologies can be used to validate a modeler's simulation? How similar does a simulation need to be to a real-world system in order for it to be effective?

NAIVE PHILOSOPHICAL APPROACHES TO EXTERNAL VALIDATION

To answer the questions raised above, Naylor and Finger (1967), in their oft-cited article, "Verification of Computer Simulation Models" describe three philosophical approaches – *empiricism*, *positive economics*, and *rationalism*. Each of these approaches or positions is designed to arm the simulation modeler with convincing evidence that a model is externally valid. Many simulation texts frequently cite these three positions as a method for teaching validation approaches (see: Emshoff and Sisson, 1970; Law and Kelton, 1991; Pegden, Shannon, and Sadowski, 1995).

EMPIRICISM

Empiricism is an objectivist stance whose believers purport that all our knowledge comes from perception. Empiricists believe that if one wants to validate a model, then one must take every construct or particular part and reduce it to something one can perceive. What one perceives is the so-called empirical foundation. Similarly, *objectivism* is a philosophy that stresses that we all have direct access to a universal set of these facts – a foundation (*foundationalism*). An empiricist believes that the observable facts are what Bertrand Russell termed the "ultimate furniture of the world" (Weinberg, 1936, p. 53). In order for an empiricist to believe that a statement made by another is true, the provider of the statement must be able to verify it by dragging the empiricist by the nose to the observable fact (*verificationism*). Anything not observable is deemed metaphysical, and therefore, meaningless.

Empiricism is an approach that requires all assumptions to be empirically tested if they are to be included in the model. No assumptions can be included in a model unless they have undergone their own empirical 'verification' process whereby they are independently tested through observable or descriptive data. The belief that propositions of a theory or parts of a model can be empirically verified in this manner is to take the position of a logical positivist, a rigorous version of empiricism.

THE PROBLEM OF INDUCTION

This position requires the reviewer to induce a model's external validity. A hierarchical, bottom-up framework of a finite set of facts and assumptions are provided to a reviewer by the simulation modeler to ‘show’ that a model is valid. However, to induce or draw a general rule or conclusion based solely on observable facts is generally agreed to be impossible. The simulation modeler is typically relying on a personal set of experiences to make a generalization about the validity of the model. Unfortunately, "no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white" (Popper, 1959, p. 27). How could one feasibly see all swans past, present, and future? It is unjustified to create universal statements or general propositions from a set of singular statements or facts; they go beyond the personal experiences of the researcher. From a purely empiricist point-of-view, one cannot appeal to the experiences of others in general because these ‘experiences’ are not in the range of experience of the other person claiming them. These other ‘experiences’ are metaphysical. Therefore, evaluative positions that rely on unqualified empiricism as a philosophical approach to immunize a simulation model from criticisms of external validity fall short on their own terms.

POSITIVE ECONOMICS

In the arguments surrounding the foundations of microeconomics, Friedman (1953) presented a version of instrumentalism that has been called positive economics. It is a positivistic response to the problem of induction. Proponents believe that scientific laws are only organizers of experience and used to make propositions about empirical observations. Scientific laws hold no other merit or value other than their use as an instrument of organization and communication.

Positive economics is the position that a model only needs to be able to predict the future with a high level of accuracy and the underlying structure or assumptions are not important. This position professes that "the truth, understandability, and rationality of the assumptions and structure [of the model] are irrelevant" (Pegden, Shannon, and Sadowski, 1995, p. 149). An example of the use of this position is in the application of the Response Surface Methodology (RSM) for objective learning validity in a simulation. RSM "is a set of statistical procedures used to develop an empirical model of the relationships between the input and output variables of a system when the inner dynamics of the system are unknown" (Carvalho, 1991). Carvalho used this technique to match learning objectives of a business course to a particular computerized business simulator – THE EXECUTIVE GAME.

RATIONALISM

Rationalists also believe that there is a foundation. Rationalist philosophers such as Descartes and Kant claimed that we all have access to some fundamental set of principles that are part of our cognitive apparatus. They believe that we do not learn from experience per se, but the innate principles in our minds organize experience for us.

Rationalism applied to simulation is the position that at least some of the underlying propositions used in a model are known to be true on a priori grounds and, therefore, do not need to be proven. Through the use of logical deduction, a verified model can then be scaffolded around these propositions.

Rationalists tend to encounter problems with rationalizing functions or mathematical thinking. Historically, rationalist principles that were thought by one generation to be intuitively obvious were thrown out by the next generation. The replacement of Euclidean geometry with non-Euclidean geometry or the replacement of particle theories with wave theories of light are examples of this practice. This happened frequently enough that rationalism as a foundation has become suspect.

RELATIVISM

Relativism emerged as a philosophy because of the inconsistencies in the aforementioned foundationalist philosophies. The failures of foundationalism, both in its empiricist and rationalist forms, led to questions of a foundation itself as a project. Relativists believe that one may not have access to any foundational principles – one's opinion is most important. It has been interpreted that when relativistic philosophers of science came upon inconsistencies in their philosophical agenda, they created conventions or rules in an effort to resolve these problems and validate their theories. In essence, they justified their decisions along utilitarian lines by explaining that they were for the good of science. Many noted philosophers of science have been read as *conventionalists* – a term usually held to be synonymous with relativism. For example, Popper, Kuhn, and Feyerabend were interpreted in this manner for their scientific methodologies.

GETTING AROUND THE EITHER/OR DICHOTOMY

If one opts for one of these positions as a method for external validation, the modeler and their models are placed in the grasp of an either/or, objectivist/relativist dichotomy: Either the model has supporting evidence that warrants its validity or the model is completely suppositional and worthless as an instrument of assessment (see: Feyerabend, 1993; Kuhn, 1962; Lakatos and Musgrave, 1970; Martin and Kleindorfer, 1991; and Popper, 1959 for an in-depth discussion of this dichotomy). Several authors have argued that simulation validity does not need to be regarded as a "dichotomous variable" (Stanislaw, 1986, p. 178). These authors argued that because simulations are never exact replicates of the system, validity is only a question regarding the degree of similarity between the model and the real-world system (Ashby, 1970; Colby, 1977; Norlen, 1975; Stanislaw, 1986). This position does not move foundationalists outside of the reaches of this problem; it merely allows them to create a convention for which they can empirically test the external validity of the simulation or its components.

THE HERMENEUTICAL POSITION

Bernstein, a strong proponent of *hermeneutics*, believes that many philosophers have been engaged in the exhaustive either/or dichotomy between objectivism and relativism to "determine the nature and scope of human rationality" (Bernstein, 1983, p. 2). On the one side sits philosophers who believe that there is a fixed foundation for our knowledge; on the other is the "intellectual and moral chaos" of relativism (Bernstein, 1983, p. 18). Bernstein argues that this dichotomy has caused great anxiety and must be exorcised from our minds. Philosophers must therefore replace this dichotomy with the therapeutic concepts of hermeneutics, praxis, and phronesis. Bernstein believes that these concepts are a different dimension, an "alternate way of thinking and of understanding" the truths of epistemology and ontology and our "being in the world" (Bernstein, 1983, pp. 118-125).

Bernstein's ideas come partially from Gadamer's (1975, 1976) notions of knowledge. Gadamer believed that knowledge lies between individual and object; its message lies in the interaction – praxis (practice). Much like a piece of art that needs a spectator to complete its understanding or meaning, knowledge involves praxis and play. It is only when spectators lose themselves in the to-and-fro motion of understanding this piece of art that play and subsequent meaning take place. Play, it seems, is at the heart of Bernstein's and Gadamer's hermeneutical philosophy.

The investigation of meaning involves a contextual and historical component with a sense of 'what is right' – phronesis. This application – phronesis or 'in council' – combined with praxis is the essential moment of the hermeneutical experience. Thus, hermeneutic philosophers such as Bernstein and Gadamer believe that meaning does not sit in the object or its interpreter, rather it is generated through an interaction that is guided by phronesis.

The way in which scientists decide what is "worthy of investigation" is based on what Bernstein and Gadamer call "effective-historical thinking" (Bernstein, 1983, p, 142). That is, people are thrown into a historical tradition that is part of them, which guides and shapes their investigations. This thinking allows for the perpetual pursuit of self-knowledge, of consciousness. This consciousness, with its effective-historical perspective, provides an ever-changing boundary of knowledge – a horizon. New ideas are learned not by quashing viewpoints and prejudice, but merging them with new viewpoints, or other horizons. The ability to merge or fuse horizons breaks the binds of Kuhn's (1962) earlier notion of incommensurability or incomparability between paradigms and enlarges and enriches consciousness. This philosophical viewpoint is termed the hermeneutical position.

Within a hermeneutical perspective, one must recognize that data are relative to specific purposes and conventions that are set in context. Simulation models must therefore be validated in much the same way as a judicial court system operates. Defendants are not proven guilty through the mere induction of empirical facts or relativistic presumptions. They are found guilty only when the prosecution has proven the case to a jury of peers 'beyond a reasonable doubt'. Sentences are then handed down based on the context and circumstance of the crime and historical cases of law. Applying the court system concept to simulation validation, Kleindorfer, O'Neil, and Ganeshan (1998, p. 1098) sum it up best:

> We must assume openness denied by both objectivists and relativists in which we can conduct meaningful dialog on a model's warrantability. This openness would imply an ability to meaningfully compare different models. The model builder or builders would be free to establish and increase the credibility of the

> model through any reasonable means. This process could involve other stakeholders, such as model users and referees of journal articles, who would share part of the responsibility of effecting model validation…Thus, the purpose should be to lend just enough structure to provide stability and lend meaning to questions of validation, yet not so much as to diminish the importance of individual freedom and ethical behavior in model validation.

JUDGMENTAL BIAS

Within this philosophy, one must be cognizant of the inclusion of judgment bias into the simulation. Irvine, Levary and McCoy (1998) state that biases can be introduced into simulation in one of four stages: "The characterization of model requirements phase, the data collection phase, the preliminary design phase, and the final design validation phase." They believe that judgmental bias can have a substantial impact on a simulation and dilute its effectiveness as a managerial or teaching tool. They classify judgmental biases into three major categories: biases related to data, biases related to decision makers, and biases related to decision makers' use of data.

Objectivists who believe that biases can be completely removed contradict the hermeneutical perspective. Bernstein believes that "there is no knowledge without preconceptions and prejudices. The task is not to remove all such preconceptions, but to test [and risk] them critically in the course of inquiry" (Bernstein, 1983, p. 128). One can never be rid of biases because in fact they are what one always starts with and help make thinking possible. In fact, Gadamer believes that "prejudices are biases to our openness to the world" (Bernstein, 1983, p. 128).

It is important that all parties are aware of the inclusion of biases and work together in an effort to understand them – not attempt to arrive at a mythical, value-free, abstract, objectivist environment. Similar to the legal instructions of a judge to a jury, parties involved in the development and evaluation of a simulation must attempt to understand their prejudices and evaluate all provided information on the merits of its ability to provide an accurate representation of the system under study.

SUMMARY

We began this paper by noting the fact that many researchers attribute the problems involved in evaluating simulations to poorly designed studies and difficulties inherent in creating a methodology of evaluation. We believe that the problem is more deep-seated than this. The literature is so cluttered with terms and concepts that it is hard to build a coherent program of validation research. A poorly designed study or ill-conceived methodology of evaluation from one perspective might be well designed and appropriate from another. For instance, we might look at simulations as experiential learning activities that allow learners to visualize situations and see the results of manipulating variables in a dynamic environment. Although simulation models need to imitate situations in such a manner that a learner can gain insight into the interaction of variables within that system, these situations do not need to be exact replicates. In fact, simplistic simulations can actually assist novice managers by focusing their attention on important variables. Thus, it might receive a very positive evaluation as a learning tool, but it might fare quite poorly as a tool for modeling actual real-world phenomena.

In order to sort out the issues, we have tried to summarize the literature in terms of the framework shown in Figure 2. That is, we have divided studies into those that address verification, internal validation, and external validation. We have viewed these in terms of three different systems -- the validation system, the development system, and the educational system. We have characterized the relationship between the validation and development systems as representational validation. We have characterized the relationship between validation and educational systems as educational validation. This brings an order and logic to the literature. More important, it heads off some of the conflicting findings and confusion resulting from seeking a common standard for evaluating simulations that have been created with divergent objectives, as illustrated by our example of a simulation that is deliberately simplistic in order to increase educational effectiveness, versus one that is necessarily complex in order to adequately represent real-world phenomena.

Finally, Table 1 presents a lexicon, summarizing key terms used in the literature on simulation validation. It anchors each of these terms in the framework of Figure 1, thus organizing our thinking for future programs of research.

REFERENCES

Anderson, P., Cannon, H.M., Malik, D. and Thavikulwat, P. (1998). Games as Instruments of Assessment," in John K. Butler, Jr., Nancy H. Leonard and Sandra W. Morgan (eds.), *Developments in Business Simulation and Experiential Learning*, V. 25, pp. 31-9.

Alessi, S.M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer Based Instruction, 15*(2), 40-47.

Ashby, W.R. (1970). Analysis of the system to be modeled, in R. M. Stogdill (Ed.), *The process of model-building in the behavioral sciences*. Columbus: Ohio State University Press.

Babbie, E. (1992). *The practice of social research (6th Ed.).* Belmont, CA: Wadsworth.

Bernstein, R.J. (1983) *Beyond objectivism and relativism: Science, hermeneutics, and praxis*. Philladelphia, PA: University of Pennsylvania Press.

Blaiwes, A.S. and Regan, J.J. (1986). Training Devices: Concepts and Progress. In J.A. Ellis (Ed.), *Military* *contributions to instructional technology*. New York, NY: Praeger Publishers.

Boocock, S.S. (1972). Validity-testing of an intergenerational relations game. *Simulation & Gaming*, 3, 29-40.

Bredemeier, M.E. and C. S. Greenblat (1981). The educational effectiveness of simulation games. *Simulation & Gaming*, 12, 307-332.

Bunker, W.M. (1978). Training effectiveness versus simulation realism. *Eleventh NTEX/Industry Conference Proceedings*. Naval Training Equipment Center: Orlando, FL.

Campbell, D.T. and Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of research and on teaching*. Chicago: Rand McNally.

Cannon, H. M. (1995). Dealing with the complexity paradox in business simulation games. *Developments in Business Simulation & Experiential Exercises*, 22, 97-103.

Carmines, E.G. and Zeller, R.A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage Publications, Inc.

Carvalho, G.F. (1991). Evaluating computerized business simulators for objective learning validity. *Simulation & Gaming*, 22(3), 328-348.

Cook, T.D. and Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Colby, K.M. (1977). On the way people and models do it. *Perspectives in Biology and Medicine*, 21, 99-104.

Cox, J.A., Wood, R., Boren, L., and Thorne, H.W. (1965). *Functional and appearance fidelity of training devices for fixed procedures tasks*. Alexandria, VA: Human Resources Research Organization. (HumRRO Tech. Rep. 65-4).

Descartes, R. (1993). *A Discourse on Method; and, Meditations on the First Philosophy* (Donald A. Cress, Trans.). Indianapolis, IN: Hackett Publishing Company, Inc. (Original work published 1641).

Dickson, John R., T. Richard Whitely, and Anthony J. Faria (1990). An empirical investigation of the internal validity of a marketing simulation game. *Developments in Business Simulations & Experiential Exercises*, 17, 47-52.

Dukes, R.L. and S. J. Waller, S.J. (1976). Toward a general evaluation model for simulation games: GEM. *Simulation & Gaming*, 7, 75-96.

Dwyer, F.M. (1974). *Strategies for improving visual learning.* State College, PA: The Pennsylvania State University, Learning Services.

Emshoff, J.R. and Sisson, R.L. (1970). *Design and use of computer simulation models*. New York, NY: The Macmillan Company.

Friedman, M. (1953). The Methodology of Positive Economics. *Essays in Positive Economics*. University of Chicago Press, Chicago, IL, 3-43.

Feyerabend, P. (1993). *Against method (3rd Ed.).* New York, NY: Verso.

Gadamer, H.G. (1975). *Truth and method*. Translated and Edited by Garrett Barden and John Cumming. New York, NY: Seabury Press.

Gadamer, H.G. (1976). *Philosophical hermeneutics*. Translated by David E. Linge. Berkeley, CA: University of California Press.

Gagne, R.M. (1954). Training devices and simulators: Some research issues. *The American Psychologist, 9*, 95-107.

Greenlaw, P.S, Herron, L.W. and Rawdon, R.H. (1962). *Business simulation in industrial and university education*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Greenlaw, P.S. and Wyman, F.P. (1973). The teaching effectiveness of games in collegiate business courses. *Simulation & Games*, 4, 259-294.

Hays, R.T. and Singer, M.J. (1989*). Simulation fidelity in training system design: Bridging the gap between reality and training*. New York, NY: Springer-Verlag.

House, W. C. and H. S. Napier (1988). A comparative study of strategic performance factors in actual and simulated business environments. *Developments in Business Simulation & Experiential Exercises*, 15, 50-55.

Irvine, S.R., Levary, R.R. and McCoy, M.S. (1998). The impact of judgmental biases on the validation of simulation models. *Simulation & gaming*, 29(2), 152-164.

Kant, I. (1996). *Critique of pure reason* (W.S. Pluhar, Trans.). Indianapolis, IN: Hackett Publishing Company, Inc. (Original work published 1781/1787).

Kibbee, J. M. (1961). Model building for management games, in A. Newgarden (ed.). *Simulation and Gaming: A Symposium*. New York: American Management Association, 8-15.

Kinkade, R. and Wheaton, G. (1972). Training Devices Design. In H. Vancoff & R. Kinkade (Eds.), *Human engineering guide to equipment design*. Washington, D.C.: American Institutes for Research.

Kleindorfer, G.B., O'Neill, L., and Ganeshan, R. (1998). Validation in simulation: Various positions in the philosophy of science. *Management Science*, 44(8), 1087-1099.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lakatos, I. and Musgrave A. (1970). *Criticism and the growth of knowledge*. Cambridge, England: Cambridge University Press.

Law, A.M. and Kelton, W.D. (1991). *Simulation modeling and analysis* (2nd Ed.). New York, NY: McGraw-Hill.

Martin, J.E., and Kleindorfer, G.B. (1991). The argumentum ad hominem and two theses about evolutionary epistemology: 'Godelian' reflections. *Metaphilosophy*, 22, 63-88.

Martin, E.L. and Waag, W.L. (1978). *Contributions of platform motion to simulator training effectiveness:*

Study I-basic contact. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Mehrez, A., A. Reichel, and R. Olami (1987). The business game versus reality: A case study in the Israeli environment. *Simulation & Games*, 18, 488-500.

Mihram, G.A. (1972). Some practical aspects of the verification and validation of simulation models. *Operations Research*, 23, 17-29.

Miles, W.G., Biggs, W.D., and Schubert, J.N. (1986). Student perceptions of skill acquisition through cases and a general management simulation: A comparison. *Simulation & Games, 17*(1), 7-24.

Miller, G.E. (1974*). Some considerations in the design and utilization of simulators for technical training.* (Report No. AFHRL-TR-74-65). Alexandria, VA.: Air Force Human Resource Laboratory.

Miller, J.G. (1978). *Living Systems*. New York, NY: McGraw Hill.

Muckler, F.A., Nygaard, J.E., O'Kelly, L.I., and Williams, A.C. (1959*). Physiological Variables in the design of flight simulators for training*. Wright-Patterson Air Force Base, OH: Wright Air Development Center.

Naylor, T.H. and Finger, J.M. (1967). Verification of computer simulation models. *Management Science*. 14, B92-B101.

Norlen, U. (1975). *Simulation model building*. New York, NY: John Wiley.

Norris, D.R. (1986). External validity of business games. *Simulation & Gaming*, 17(4), 447-459.

Norris, D.R. and C. A. Snyder (1982). External validation of simulation games. *Simulation & Gaming*, 13, 73-85.

Parasuraman, A. (1981). Assessing the worth of business simulation games. *Simulation & Games*, 12:2, 189-200.

Pegden, C.D., Shannon, R.E. and Sadowski, R.P. (1995). *Introduction to simulation using SIMAN. (2nd Ed.).* Hightstown, NJ: McGraw-Hill, Inc.

Peters, V., Vissers, G. and Heijne, G. (1998). The validity of games. *Simulation & Gaming,* 29(1), 20-30.

Pierfy, D. (1977). Comparative simulation game research: Stumbling bocks and stepping stones. *Simulation & Games*, 8, 255-268.

Popper, K.R. (1959). The logic of scientific discovery. New York, NY: Routledge.

Stanislaw, H. (1986). Tests of computer simulation validity: What do they measure? *Simulation & Gaming*, 17(2), 173-191.

Vandierendock, A. (1975). *Inferential simulation: Hypothesis-testing by computer simulation*. Nederlands Tijdschrift voor de Psychologie (Amsterdam), 30, 677-700.

Weinberg, J.R. (1936). *An examination of logical positivism*. New York, NY: Harcourt, Brace and Company.

Wolfe, J. (1981). Research on the learning effectiveness of business simulation games: A review of the state of the science. *ABSEL Proceedings*, 72.

Wolfe, J. (1985). The teaching effectiveness of games in collegiate business courses: A 1973-1983 update. *Simulation & Games*, 16(3), 251-288.

Wolfe, J. (1990). The guide to experiential learning, in J.W. Gentry (Ed.). *The evaluation of computer-based business games: Methodology, findings, and future needs*. New York, NY: Nichols.

Wolfe, J. and Crookall, D. (1998). Developing a scientific knowledge of simulation/gaming. *Simulation & Gaming, 29*(1), 7-19.

Wolfe, J, and R. Jackson (1989). An Investigation of the Need for Algorithmic Validity. *Simulation & Games* 20:3 (September), 272-291.

Wolfe, J. and C. R. Roberts (1986). The external validity of a business management game: A five-year longitudinal study. *Simulation & Games*, 17, 45-49.

Wolfe, J. and C. R. Roberts (1993). A further study of the external validity of business games: Five year peer group indicators. *Simulation & Gaming*, 24(1), 21-33.