

# Constructs of simulation evaluation

**Andrew Hale Feinstein**

*University of Nevada, Las Vegas*

**Hugh M. Cannon**

*Wayne State University*

*Although instructional research on simulation has been around for almost 40 years, validation research has failed to hold itself to a common, scientifically 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, a failure to adhere to a generally accepted research taxonomy, and no well-defined constructs with which to assess learning outcomes. This article seeks to address the problem by reviewing the various concepts employed in the literature of simulation evaluation and integrating them into a coherent framework from which the evaluative process may proceed in a more systematic manner.*

**KEYWORDS:** *simulation assessment; simulation evaluation; simulation fidelity; simulation validity; simulation verifiability.*

One of the major problems of simulations is how to “evaluate the training effectiveness [of a simulation]” (Hays & Singer, 1989, p. 193). Although for more than 40 years, researchers have lauded the benefits of simulation (Wolfe & Crookall, 1998), very few of these claims are supported with substantial research (Butler, Markulis, & Strang, 1988; Miles, Biggs, & Schubert, 1986; Wolfe, 1981, 1985, 1990; Wolfe & Crookall, 1998). In the words of Wolfe and Crookall (1998),

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. (p. 8)

Many of the researchers cited above attribute the relative lack of progress in simulation evaluation to poorly designed studies and the difficulties inherent in creating an acceptable methodology of evaluation. But difficult or not, to progress as a discipline, simulation and gaming must develop and adhere to such a methodology or, as the quote from Wolfe and Crookall (1998) implies, a set of methodologies that are appropriate to the different types of simulation/gaming phenomena being considered. This, in turn, requires a clear definition of the relevant constructs—the “vocabulary” with which we might discuss the evaluative process.

The purpose of this article is to review the literature on simulation evaluation, summarize the key concepts appearing in the literature, and develop a coherent framework for pursuing the evaluation problem. In doing so, we will address three major constructs: *fidelity*, *verification*, and *validation*. Whereas all three are prominent in the literature, we will argue that, at least as they relate to simulation and gaming as a tool for training and education, validation is the most important. Indeed, focusing on the issue of fidelity can actually detract from the development of educationally effective games. Whereas verification is important, it is already relatively well understood by game developers, and as with fidelity, focusing on it can actually get in the way of validation. Indeed, verification has often been implicitly substituted for validation by many game developers, distracting them from the thornier problems of validation.

### 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 & Singer, 1989, p. 1).

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

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. (p. 50)

### 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 1970s (Blaiwes & Regan, 1986; Bunker, 1978; Cox, Wood, Boren, & Thorne, 1965; Greenlaw, Herron, & Rawdon, 1962; Kibbee, 1961; Kinkade & Wheaton, 1972; Martin & Waag, 1978; J. G. Miller, 1978; Muckler, Nygaard, O'Kelly, & Williams, 1959). 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 (Alessi, 1988; Dwyer, 1974; Gagne, 1954; G. E. Miller, 1974). Furthermore, Martin and Waag (1978) determined that high fidelity can actually hinder effective training and learning because it overstimulates novice trainees.

Blaiwes and Regan (1986) believed that in simulation, the goal is to provide a learning environment, not a vehicle for trainees to exhibit perfect performance. Kibbee (1961) believed 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 on for creating effective learning environments.

Hays and Singer (1989) pointed out that it can be cost-effective for novice trainees to utilize low-fidelity devices during the first stages of learning. These authors also believed that a simulator does not need to be an exact representation of the real world to provide effective training. In fact, they felt that it might be necessary to "depart from realism in order to provide the most effective training" (p. 15). They also believed 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" (p. 15).

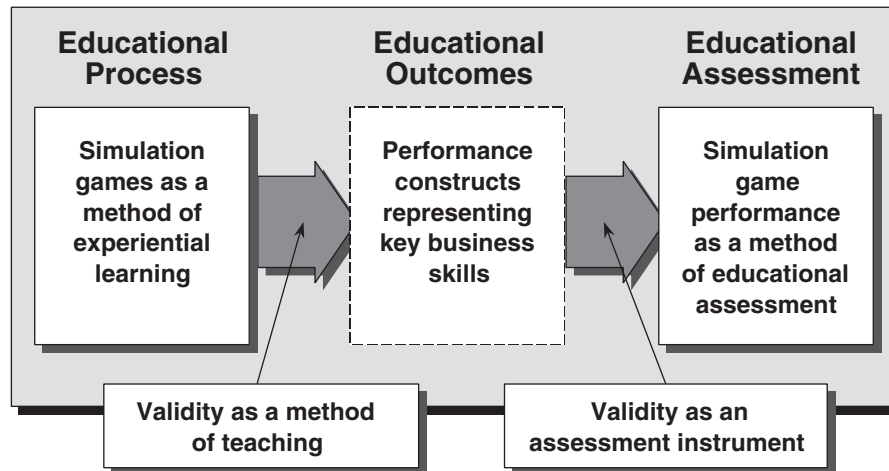
From the perspective of management education, Cannon (1995) argued that the apparent inverse relationship between fidelity and learning parallels what we find in the real world. The workings of human organizational and environmental systems are so inherently complex that no one is capable of dealing with them effectively. As a result, people develop techniques for simplifying the decisions they have to make. Simulation games can model the same techniques, in this case adding an element of fidelity to gaming process. However, the slavish pursuit of fidelity can be devastating to the learning process.

### Simulation verification

The general notion of validation incorporates two different processes: verification and validation (Pegden, Shannon, & 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, & 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 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



**FIGURE 1: Two Faces of Simulation Game Validation**  
 SOURCE: Anderson, Cannon, Malik, and Thavikulwat (1998).

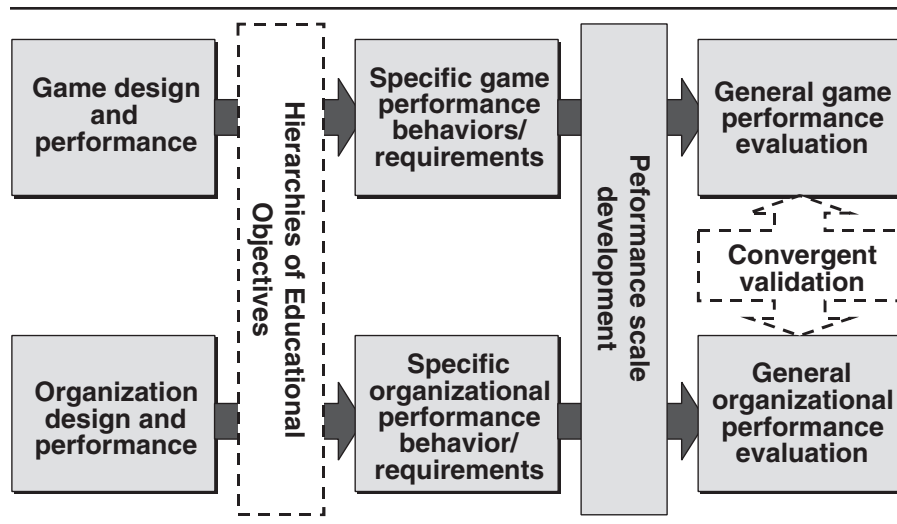
eliminating errors. Typically, this testing is done first with alpha tests by the simulation developer and, later, with beta tests, where the simulation is run in a variety of conditions by independent users.

All of this is well known and appreciated by most simulation game developers. What is less well known, and certainly less appreciated, is that verification is only a necessary, not a sufficient, condition for validity. Too often, developers produce what appear to be brilliant models but have no idea of their educational effect, hence, their validity. In this sense, verification can be a trap. It can distract developers from key issues of validation, notwithstanding its critical status as a necessary condition for validity.

### The problem of validation: An overview

During the decade of the 1990s, the Association for Business Simulations and Experiential Learning (ABSEL) Assessment Committee was 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, and the committee's efforts illustrate the validation issues we propose to address. Figure 1 provides an overall picture of the problems the committee faced.

Figure 1 suggests that simulation games might serve one of two major purposes relative to business management. First, using a game as a method of experiential learning in business suggests that its validity will depend on how well it prepares students to understand, select, and appropriately use a set of key business skills. Second, it



**FIGURE 2: Evaluating Simulation Performance Criteria**  
SOURCE: Cannon and Burns (1999).

suggests that regardless of any educational efficacy, it might be used to assess the skills that businesspeople already possess.

In service of these both these perspectives, Cannon and Burns (1999) have suggested a method for operationalizing the business skills that represent a common ground between the two simulation game applications. Drawing on the work of Kibler, Cegala, Barker, and Miles (1974), who in turn drew on the work of Bloom, Englehart, Furst, Hill, and Krathwohl (1956); Krathwohl, Bloom, and Masia (1964); and Simpson (1974), they suggested a three-dimensional taxonomy for developing scales to measure the performance constructs representing the skills.

If the scales developed through Cannon and Burns's (1999) process do indeed represent valid educational objectives, and if they indeed reflect skills that are both developed and exhibited in the course of playing a business simulation game, the game can be used as both a learning tool and an instrument to assess educational accomplishment or general employee competence. That is, player performance can be either a dependent or an independent variable, either a consequence or a predictor of player competence.

Cannon and Burns (1999) went on to describe the method by which their model would be validated. This is portrayed in Figure 2. It involves two parallel processes, first, evaluating the actual job performance requirements of a target organization or set of organizational positions, simultaneously designing a simulation game to represent the same organization requirements. The cognitive, affective, and psychomotor taxonomies (hierarchies) of educational objectives provide a conceptual lens through which to evaluate the skills (performance behaviors) required by both the game and the organization. Once conceptualized, these behaviors would be reduced to items on a set of evaluative scales, to be measured by any number of different possible mechanisms. For

instance, a performance behavior might be the ability to forecast intermediate results or anticipate the immediate consequences of a specific managerial decision (Teach, 1987). These could be measured by observation, or perhaps by requiring game players to submit forecasts prior to determining the actual consequences of their decisions.

Once the scales have been developed, they can be validated by separating a priori the most and least successful performers, both in the simulation game and in the actual organization. The items can then be evaluated to determine how well they discriminate between the two groups. Finally, these results can then be compared for the simulation players versus the organizational participants to see whether the same items discriminate in each of the two situations, providing a kind of *convergent* validity.

### A lexicon of simulation validation research

Cannon and Burns's (1999) framework represents only one of a host of proposed approaches for simulation evaluation. However, by describing it in some detail, we can use it as a metaphor for the validation process in general, thus providing a relatively concrete vehicle for our discussion.

Prior to actually discussing the general process(es) of validation, our task will be to review the broader literature and identify the various terms used in validation research. Our discussion will then proceed to sort them into a coherent, logical framework. Table 1 presents a summary of the terminology.

Notwithstanding the plethora of terms, their meanings tend to be related. Indeed, they can be roughly understood in terms of two basic dimensions: game development versus application, and internal versus external validity. These are reflected in the framework shown in Figure 3. The *developmental system* represents issues regarding the actual development of a simulation game, drawing on principles of *representational validity*. The *educational system* represents issues involving the learning process, as the game is actually applied in a teaching environment, drawing on principles of *educational validity*. *Internal validity*, roughly speaking, addresses the extent to which a simulation functions in the intended manner. *External validity* asks whether this internal functioning corresponds to relevant phenomena outside the simulation.

#### Representational validity

Given our earlier discussion of fidelity, one might be tempted to downplay the importance of representational validation. After all, we are interested in educational impact, not fidelity. And as we have seen, fidelity can actually get in the way of educational validity. To do so, however, would be a mistake. We need only look at the terms summarized in Table 1 to see that most of them deal with representational, not educational, validity. However, a slavish dedication to fidelity would be equally ill advised. The secret to resolving the apparent paradox is recognizing the fact that representational validity addresses a process, not an end result. What Figure 3 calls the *developmental system* portrays the process of identifying and abstracting appropriate concepts

**TABLE 1: Concepts Related to Simulation Validation Research**

<i>Concept</i>	<i>Definition</i>	<i>Reference</i>
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 validity	Does the model adequately represent the real-world system? (Special case of external representational validity. A key issue in establishing internal educational validity. May be seen as a more abstract version of accuracy.)	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.)	Babbie, 1992, pp. 132-133; Carmines and Zeller, 1979
Content validity	How complete is the simulation model, relative to the demands imposed by the purpose for which the model was developed? (A special case of internal representational validity.)	Babbie, 1992, pp. 132-133; Carmines and Zeller, 1979
Convergent validity	How well do simulation performance results compare with other measures of comparable competencies? (A measure of external validity.)	Cannon and Burns, 1999.
Criterion (predictive) validity	Does the model effectively predict real-world situations? (Special case of external representational validity. See accuracy, realism, empirical validity.)	Babbie, 1992; Carmines and Zeller, 1979
Educational validity	Does the simulation provide a valid learning experience and/or assessment of learning? (As contrasted with representational validity.)	Conceptualized in this article
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
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
Internal validity	Do a model's relationships represent true causality? (An issue relating to external representational validity.)	Cook and Campbell, 1979

*(continued)*

TABLE 1 (continued)

<i>Concept</i>	<i>Definition</i>	<i>Reference</i>
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
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
Representational validity	Does the simulation provide a valid representation of a desired phenomenon?	Conceptualized in this article
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 (face validity)	Does the simulation model appear to represent real-life phenomena? Often referred to as face validity. (Reflection of perceived external representational validity.)	Kibbee, 1961

to be modeled, actually modeling them in the context of game structure and logic, and then designing them into an actual simulation game.

From this perspective, we find terms such as *accuracy* (Dukes & Waller, 1976), *empirical validity* (Boocock, 1972), *event validity* (Mihram, 1972), and *realism* (Norris, 1986) generally less useful than *conceptual validity* (Pegden, Shannon, & Sadowski, 1995). Indeed, conceptual validity is essentially the same as accuracy, except that it deals with the theoretical rather than literal essence of the phenomena being modeled. To be conceptually valid, the simulation model need only contain a degree of "homomorphism" (Stanislaw, 1986, p. 177; Vandierendock, 1975) between itself and the system it is modeling, commensurate with a set of objectives. In our view, the difference is analogous to the way a great piece of art captures the essence of reality better than an "accurate" photograph. In art, the artist harnesses the medium to capture what he or she considers the key aspects of reality, whereas in a simple photograph, the definition of reality is determined by the mechanics of the equipment, with no conscious interpretative intervention. In the world of simulation games, the artistic interpretation is called theory, the modeler's conceptualization of the phenomena being modeled.



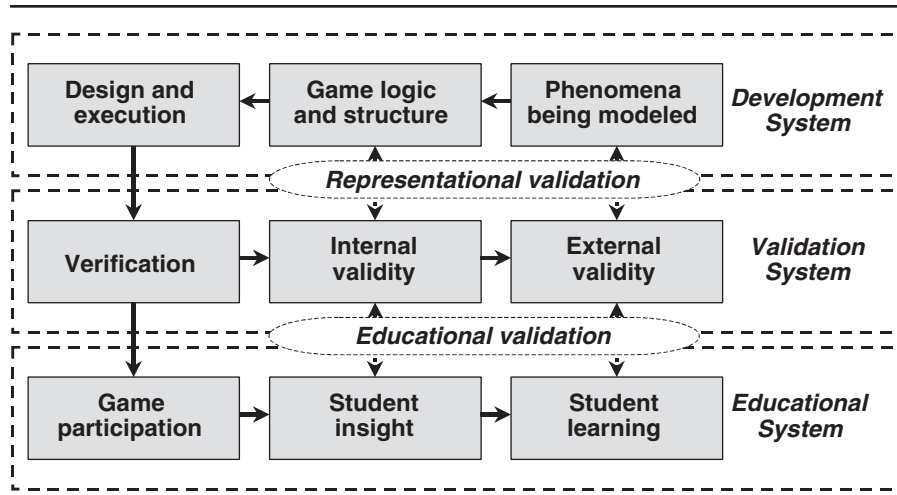


FIGURE 3: Two Patterns of Simulation Game Validation

Continuing, the appropriateness of the modeler's selection of phenomena is called *content validity*, whereas the correctness with which the model is constructed is called *construct validity* (Babbie, 1992, pp. 132-133; Carmines & Zeller, 1979). Content validity, then, represents the artistic judgment that defines the modeler's picture of the simulated reality. The picture itself, however, is a function not only of the selection, but how well the phenomena are modeled. If an accurate model is like a photograph, *algorithmic validity* (Wolfe & Jackson, 1989) is the arbiter of accuracy. But if we view the simulation as an artistic creation, we are seeking conceptual validity rather than simple accuracy. For conceptual validity, the arbiter is construct validity (Babbie, 1992, pp. 132-133; Carmines & Zeller, 1979). That is, do the various concepts represented by variables in the model fit together correctly. This is analogous to algorithmic validity, but algorithmic validity applies only to situations involving a high degree of "instrumental perfection," where the relationship between variables can be specified by clear, commonly accepted mathematical relationships. Construct validity implies that the relationships between variables are correct, but they can be more subjective and modeled by any number of heuristic devices. One measure of conceptual validity—the artistic touch, if you will—is the degree to which this selection "rings true" with the simulation user or evaluator. This is *believability* (Pegden, Shannon, & Sadowski, 1995), *plausibility* (Boocock, 1972), *verisimilitude* (Kibbee, 1961), or what we commonly refer to as *face validity*.

Note that the process we have just described, the identification and interpretation of relevant phenomena, is exactly what Cannon and Burns (1999) presented (see Figure 2). As a means of developing conceptual validity, they used the lens of the three taxonomies of educational objectives. Whereas effective business skills might look very different in different settings, their essence is defined by the cognitive, affective, and psychomotor patterns that define them. The development of scales to measure

these skills suggests a set of scale-development techniques that incorporate various measures of construct validity. Extensions of these techniques, such as modern methods of path analysis and structural modeling, lend themselves to establishing the construct validity of the simulation model itself (Loehlin, 1998). Testing measured simulation performance to see whether it is higher for students and managers who are recognized high performers provides criterion validity (Babbie, 1992; Carmines & Zeller, 1979), which gives us confidence that the skills being measured are indeed valid. Establishing convergent validity by demonstrating the fact that the measures work in both a simulated and organizational environment provides a measure of external validity (Cook & Campbell, 1979), suggesting that the results are not an artifact of the gaming environment itself. Ensuring that the skills utilized in the simulation are the same as those used in the kind of organization it is intended to simulate is what Pegden, Shannon, and Sadowski (1995) called *operational validity*.

### **Educational validity**

The payoff of establishing representational validity, in the artistic sense we have discussed, should be in the area of educational validity. That is, a well-developed simulation game, if designed for educational purposes, should serve those purposes. Educational validation seeks to determine whether this is the case.

Here we can again draw on Cannon and Burns's (1999) model as an example. In the second stage of their model, they sought to evaluate the design and performance of a game against the structure posited by educational taxonomies. The taxonomies propose patterns of thinking, feeling, and acting—student insights—that the game should stimulate as part of the learning experience. To the extent that these can be observed, the educational aspect of the simulation could be said to have conceptual validity. Insofar as these insights form in the patterns and sequences posited in the development system of Figure 3, the educational aspect of the game could be said to have construct validity. Moving on in the educational system, one of the purposes for creating scales to measure business skills is to see whether students actually acquire the targeted skills (criterion validity). The fact that the skills have been validated against those required in an actual business organization provides a measure of external validity.

We see, then, that whereas most of the formal discussions of validity tend to address the representational domain, the process of educational validation shown in Figure 3 parallels that of representational validation. The difference, as we have noted earlier, is that educational validation comes at a later application stage of the game validation process. What is being evaluated in this stage is not the structure of the game but, rather, the effects it has on those who play it.

### **Internal validity**

Now, having distinguished between development and application, we can look at the second dimension by which the validation process is typically classified. Campbell (Campbell & Stanley, 1963; Cook & Campbell, 1979) has divided the general concept

into two types of validity: internal and external. Cook and Campbell (1979) described 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. (p. 37)

As we have just seen, Figure 3 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) called *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 is discussed as content validity, and how correctly the variables are related to each other to form strategic and environmental constructs, that is, construct validity (Babbie, 1992, pp. 132-133; Carmines & Zeller, 1979).

The second form of internal validity addresses the degree to which game participants understand the game and play it with insight, following what we have referred 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. 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. In terms of the Cannon and Burns (1999) model (see Figure 2), students would have to see themselves as needing, learning, and using the kinds of skills being conceptualized in the scale-development process.

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" vs. "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 & Greenblat, 1981; Greenlaw & 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.

They are learning to forecast, make pricing decisions, conceptualize and evaluate alternative strategies, and so forth. However, it is not necessarily true for representational validity. Indeed, as we have seen, a simulation that faithfully represents strategic cause and effect might well be so complex that students never see the relationships (Cannon, 1995). A realistic market forecast may involve such a complex range of competitive and environmental interactions that students would simply see changes in sales as a random event.

Nor is internal validity an absolute assurance of educational effectiveness, even relative to educational validity. We are assuming that the artistic modeler created a simulation that faithfully (i.e., with strong internal validity) stimulates a set of educationally desirable responses on the part of students. However, this is not necessarily the case. For instance, in the interest of teaching the effect of advertising in consumer markets, a game might emphasize the advertising function and end up teaching students that advertising is always the primary key to marketing success. The game would be internally valid, but externally disastrous!

### External validity

Figures 2 and 3 both suggest that the logic and structure of a simulation game are reflections of some real-world phenomenon. This is a manifestation of external validity (Boocock, 1972; Dukes & Waller, 1976; Mihram, 1972; Norris & Snyder, 1982; Wolfe & Roberts, 1986, 1993). As we see from Figure 3, the cycle of simulation game development begins with questions involving external validity—what are the most relevant phenomena to be modeled, the answer being manifest primarily in the form of content and conceptual validity. Once decided, the task is to ensure that the game is internally consistent. Next, the validation questions whether the game actually functions as intended. And finally, it questions whether the game's function is really as useful as the original design hoped it would be.

Whereas most researchers tend to agree on the nature of internal validity, there is considerable disagreement regarding external validity. This, of course, is what we would expect, given the artistic element that goes into determining what is relevant in the external environment and worth modeling. However, we believe that much of the problem can be attributed to the fact that the validity of the simulation is keyed to its objectives. This is highlighted in Figure 3 by the distinction between representational and educational validation. For instance, Mehrez, Reichel, and Olami (1987) and House and Napier (1988) have studied 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's (1986) realism, and criterion (predictive) validity (Babbie, 1992; Carmines & Zeller, 1979). 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. Although 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 to perform well in a business simulation game (validation as an assessment instrument).

We have argued that true representational validity would be based on content and conceptual validity, determined by an artistic theory-based modeler. Nevertheless, the tendency in the representational validation stage is to focus too much on the real world, to make games too complex, too accurate. This puts a special burden on *external educational validation*. Here, the question is how well does the educational process actually work in teaching real-world skills, a question that featured prominently in Cannon and Burns (1999)? Although perceptions generally take a back seat to reality in scientific endeavor, here is a place where theorists speak of verisimilitude (Kibbee, 1961), plausibility (Boocock, 1972), and believability (Pegden, Shannon, & Sadowski, 1995)—the perception of validity—as important performance criteria. These terms do not directly represent any form of external validity but only the perception of it. Rather, they are indicators (and stimulants of) motivation and insight—both of which are related most directly on internal validity. But here, the perception tends to become a self-fulfilling prophecy. The motivation and insight stimulate students to learn, which, if the game is even loosely productive of relevant business skills, tends to increase the level of external validity.

### Summary

We began this article 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.

To sort out the issues, we have tried to summarize the literature in terms of the framework shown in Table 1. That is, we have divided studies into those that address verification, internal validation, and external validation, with special emphasis on the

latter two. We have then 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 to increase educational effectiveness versus one that is necessarily complex 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 3, thus organizing our thinking for future programs of research.

## References

- Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer Based Instruction, 15*(2), 40-47.
- Anderson, P., Cannon, H. M., Malik, D., & Thavikulwat, P. (1998). Games as instruments of assessment. In J. K. Butler Jr., N. H. Leonard, & S. W. Morgan (Eds.), *Developments in business simulation and experiential learning* (Vol. 25, pp. 31-39). Association for Business Simulation and Experiential Learning/Savant Electronic Publishing, Inc. Available from <http://www.absel.org>
- Babbie, E. (1992). *The practice of social research* (6th ed.). Belmont, CA: Wadsworth.
- Blaiwes, A. S., & Regan, J. J. (1986). Training devices: Concepts and progress. In J. A. Ellis (Ed.), *Military contributions to instructional technology*. New York: Praeger.
- Bloom, B. S., Englehart, N. D., Furst, E. J., Hill, W., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives—The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay.
- Boocock, S. S. (1972). Validity-testing of an intergenerational relations game. *Simulation & Games, 3*, 29-40.
- Bredemeier, M. E., & Greenblat, C. S. (1981). The educational effectiveness of simulation games. *Simulation & Games, 12*, 307-332.
- Bunker, W. M. (1978). Training effectiveness versus simulation realism. In *Eleventh NTEX/Industry Conference Proceedings*. Orlando, FL: Naval Training Equipment Center.
- Butler, R. J., Markulis, P. M., & Strang, D. R. (1988). Where are we? An analysis of the methods and focus of the research on simulation gaming. *Simulation & Games, 19*, 3-26.
- Campbell, D. T., & 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.
- Cannon, H. M., & Burns, A. C. (1999). A framework for assessing the competencies reflected in simulation performance. *Developments in Business Simulation & Experiential Exercises, 26*, 40-44.
- Carmine, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cox, J. A., Wood, R., Boren, L., & Thorne, H. W. (1965). *Functional and appearance fidelity of training devices for fixed procedures tasks* (HumRRO Tech. Rep. No. 65-4). Alexandria, VA: Human Resources Research Organization.

- Dickson, J. R., Whitely, R. R., & Faria, A. J. (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., & Waller, S. J. (1976). Toward a general evaluation model for simulation games: GEM. *Simulation & Games, 7*, 75-96.
- Dwyer, F. M. (1974). *Strategies for improving visual learning*. University Park: Pennsylvania State University, Learning Services.
- Gagne, R. M. (1954). Training devices and simulators: Some research issues. *American Psychologist, 9*, 95-107.
- Greenlaw, P. S., Herron, L. W., & Rawdon, R. H. (1962). *Business simulation in industrial and university education*. Englewood Cliffs, NJ: Prentice Hall.
- Greenlaw, P. S., & Wyman, F. P. (1973). The teaching effectiveness of games in collegiate business courses. *Simulation & Games, 4*, 259-294.
- Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training*. New York: Springer-Verlag.
- House, W. C., & Napier, H. S. (1988). A comparative study of strategic performance factors in actual and simulated business environments. *Developments in Business Simulation & Experiential Exercises, 15*, 50-55.
- Kibbee, J. M. (1961). Model building for management games. In A. Newgarden (Ed.), *Simulation and gaming: A symposium* (pp. 8-15). New York: American Management Association.
- Kibler, R. S., Cegala, D. J., Barker, L. L., & Miles, D. T. (1974). *Objectives for instruction and evaluation*. Boston: Allyn & Bacon.
- Kinkade, R., & Wheaton, G. (1972). Training devices design. In H. Vancoff & R. Kinkade (Eds.), *Human engineering guide to equipment design*. Washington, DC: American Institutes for Research.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives—The classification of educational goals. Handbook II: Affective domain*. New York: David McKay.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Martin, E. L., & 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., Reichel, A., & Olami, R. (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., & Schubert, J. N. (1986). Student perceptions of skill acquisition through cases and a general management simulation: A comparison. *Simulation & Games, 17*, 7-24.
- Miller, G. E. (1974). *Some considerations in the design and utilization of simulators for technical training* (Rep. No. AFHRL-TR-74-65). Alexandria, VA: Air Force Human Resource Laboratory.
- Miller, J. G. (1978). *Living systems*. New York: McGraw-Hill.
- Muckler, F. A., Nygaard, J. E., O'Kelly, L. I., & Williams, A. C. (1959). *Physiological variables in the design of flight simulators for training*. Wright-Patterson Air Force Base, OH: Wright Air Development Center.
- Norris, D. R. (1986). External validity of business games. *Simulation & Games, 17*, 447-459.
- Norris, D. R., & Snyder, C. A. (1982). External validation of simulation games. *Simulation & Games, 13*, 73-85.
- Parasuraman, A. (1981). Assessing the worth of business simulation games. *Simulation & Games, 12*, 189-200.
- Pegden, C. D., Shannon, R. E., & Sadowski, R. P. (1995). *Introduction to simulation using SIMAN* (2nd ed.). Hightstown, NJ: McGraw-Hill.
- Pierfy, D. (1977). Comparative simulation game research: Stumbling blocks and stepping stones. *Simulation & Games, 8*, 255-268.
- Simpson, E. J. (1974). The classification of educational objectives in the psychomotor domain. In R. S. Kibler, D. J. Cegala, L. L. Barker, & D. T. Miles (Eds.), *Objectives for instruction and evaluation*. Boston: Allyn & Bacon.

- Stanislaw, H. (1986). Tests of computer simulation validity: What do they measure? *Simulation & Games*, 17, 173-191.
- Teach, R. D. (1987). Profits: The false prophet. *Developments in Business Simulation & Experiential Exercises*, 14, 205-207.
- Vandierendock, A. (1975). Inferential simulation: Hypothesis-testing by computer simulation. *Nederlands Tijdschrift voor de Psychologie*, 30, 677-700.
- Wolfe, J. (1981). Research on the learning effectiveness of business simulation games: A review of the state of the science. *Developments in Business Simulation & Experiential Exercises*, 8, 72.
- Wolfe, J. (1985). The teaching effectiveness of games in collegiate business courses: A 1973-1983 update. *Simulation & Games*, 16, 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: Nichols.
- Wolfe, J., & Crookall, D. (1998). Developing a scientific knowledge of simulation/gaming. *Simulation & Gaming*, 29, 7-19.
- Wolfe, J., & Jackson, R. (1989). An investigation of the need for algorithmic validity. *Simulation & Games*, 20, 272-291.
- Wolfe, J., & Roberts, C. R. (1986). The external validity of a business management game: A five-year longitudinal study. *Simulation & Games*, 17, 45-49.
- Wolfe, J., & Roberts, C. R. (1993). A further study of the external validity of business games: Five year peer group indicators. *Simulation & Gaming*, 24, 21-33.

*Andrew Hale Feinstein is an assistant professor of food and beverage management at the University of Nevada, Las Vegas. He holds a PhD from Pennsylvania State University, where he was an academic computing fellow. He also holds an MS and a BS in hotel administration from the University of Nevada, Las Vegas.*

*Hugh M. Cannon is the Adcraft/Simons-Michelson Professor of Marketing at Wayne State University. He holds an MBA and a PhD from New York University and an AB from Brigham Young University. He is a frequent contributor to the advertising and marketing literature, having published and/or presented more than 150 articles in scholarly journals, books, or conferences. He has served as executive vice president of the Association for Business Simulations and Experiential Learning (ABSEL) for many years and is a fellow in the organization.*

ADDRESSES: AHF: Department of Food and Beverage Management, UNLV College of Hotel Administration, 4505 Maryland Parkway Box 456022, Las Vegas, NV 89154; telephone: +1 702-895-1795; fax: +1 702-895-4871; e-mail: andyf@unlv.edu. HMC: Adcraft/Simons-Michelson Professor of Marketing, Wayne State University, 5021 Cass Avenue, Suite 300, Detroit, MI 48202-3930; telephone: +1 313-577-4551; fax: +1 313-577-5486; e-mail: hugh.cannon@wayne.edu.