# A review of scholarship on assessing experiential learning effectiveness

**Jerry Gosen**
**John Washbush**
*University Wisconsin, Whitewater*

*This article focuses on the research and scholarship dealing with the assessment of the experiential approach in both its computer-based and human-based forms. It covers two forms of assessment. First it covers explicit attempts to assess the validity of experiential learning, and second it focuses on measuring the effectiveness of the experiential approach. The authors classify the literature on both validity and measurement into critiques and prescriptions for ideal research on one hand and empirical research studies on the other. The empirical research reviewed supports the notion that experiential learning is effective. However, the studies showing these results reflect a long-standing trend of not meeting the highest of research design and measurement standards. Thus, the authors believe any conclusion about the effectiveness of these teaching approaches must be tentative.*

KEYWORDS: *assessment; computer-based business simulations; experiential learning; measurement; validity*

This article focuses on the research and scholarship dealing with the assessment of learning in classrooms and other learning environments using the experiential approach in both its computer-based and human-based forms. This literature includes empirical research, critiques of this research, and prescriptions for how to improve whatever research is undertaken.

Assessing outcomes in learning environments is very important, and we believe it is possible to successfully assess experiential learning environments. Outside of a very few studies, however, this has not yet been consistently accomplished. We understand the problems and pitfalls associated with undertaking such research, and we will present a seemingly rational understanding for why the evaluation ideal has been so elusive. We will also recommend solutions that are difficult but not impossible to implement. Through our review it is hoped the field will critically evaluate what it has done in the past and what it must do in the future if experiential learning is to reach its fullest potential.

SIMULATION & GAMING, Vol. 35 No. 2, June 2004 270-293
DOI: 10.1177/1046878104263544

## Why assessment is important

Assessment is a necessary complement to purpose. Perhaps the most obvious example is the business world, which assesses effectiveness with bottom-line results. The effectiveness of an army or football team is assessed by whether or not it wins battles or games, and a police department's effectiveness is assessed by its crime-solving rate. Similarly, it seems reasonable and appropriate for learning institutions to determine whether or not their students actually learn. Reviews of Association for Business Simulation and Experiential Learning (ABSEL) proceedings and issues of *Simulation & Gaming* (Anderson & Lawton, 1997; Butler, Markulus, & Strang, 1985) suggest that many scholars in the fields of simulations and experiential learning are indeed concerned with whether or not learning occurs with these approaches.

However, learning is a complex construct, hard to pin down and therefore difficult to measure. In addition, according to J. W. Gentry, Commuri, Burns, and Dickenson (1998), the failings of pedagogical assessment research are well documented. Thus, it has occurred to some (e.g., Faria & Nulsen, 1978; Schreier, 1981) to forget about attempting to carefully assess effectiveness.

However, the educator is plagued with opportunity-cost and time-use choices (Burns, Gentry, & Wolfe, 1990) as well as effectiveness concerns. Furthermore, it makes sense to try to discover which teaching approaches work and which ones do not, and that certainly is true for simulation and experiential learning approaches. As a field, we want to know whether the approaches we use are effective or not, and knowledge about effectiveness is best attained via objective and rigorous research.

## Categorizing past research

In a review of 25 years of ABSEL research on business simulations, Faria (2001) identified the following three focal research areas: correlates of simulation performance, the effectiveness of games in strategic management courses, and what business games teach. The last two of these have been accompanied by an increasing concern for the validity of the learning environments created and the measurement of what is occurring in those environments.

We will categorize the literature in a slightly different manner. We focus first on the correlates of performance, second on the validity of simulations, and third on the instruments used to measure the effectiveness of simulations and experiential exercises as teaching/learning systems. This last area deals with the field's measurement issues. As constructs, validity and measurement issues usually overlap. In this article, we separate these two because first, in some validity studies, measurement has not been important and second, in many of the studies in which measurement was very important, the focus on validity was implicit and seemingly minor. We consider research on the effectiveness of games in strategic management courses to be a subset of both validity studies and measurement issue studies—validity studies at least in part because the purpose of these studies was to investigate whether simulations were valid and measurement studies because these strategic management investigations produced instruments intended to assess simulation effectiveness.

Having reviewed a large number of studies and articles, Faria (2001) concluded that although still present, interest in correlates of performance was waning compared to interest in what simulations have to teach. The early interest in simulation performance was quite reasonable. By providing dynamic environments for competition in business-like situations, simulations could be expected to enable and invite participants to analyze, synthesize, evaluate, and apply knowledge—that is, to use higher level cognitive skills (Bloom, Englehart, First, Hill, & Krathwohl, 1956) to compete successfully. It was assumed successful players would make better use of these skills than less successful ones. Thus, it seemed reasonable simulation performance might be a proxy for learning.

However, there was also an interest among simulation scholars to justify the simulation as a valid teaching approach regardless of how well teams performed. This was in part because the effectiveness of the simulation approach was challenged by those who used traditional approaches, such as the lecture-discussion and case approaches. To justify value, it seemed reasonable learning would result from the exercise of the aforementioned higher level cognitive skills. In addition, by adding competitive environments, it seemed reasonable participants should exhibit positive affect (Krathwohl, Bloom, & Masia, 1964) and thus involvement and positive attitudes should be fostered by the learning environments created. It was therefore appropriate for simulation users to provide evidence of the potential educational value of this new methodology and to try to identify what cognitive abilities, personality factors, and behavioral actions could provide both optimal learning environments and superior performance.

However, although it seemed appropriate to undertake research assessing the value of simulations, the lion's share of early studies have focused on performance in the simulation. On the cognitive side, factors that might reflect knowledge, ability, and skills fell under the research microscope that included aptitude scores in the form of ACTs and SATs, grades, major, and other measures of academic abilities and experience. Personality-related and behavioral factors have included locus of control, decision and learning styles, gender, motivation, communication, decision processes, goal setting, motivation, cohesion, team building, functional and strategic management skills, attitudes, planning, and interest. In addition, a variety of situational factors such as team size, grading and grade weighting, number of periods of play, degree of instructor explanation and involvement, simulation complexity, and types of other course or program activities have been explored. Predictably, results have been mixed probably because of the diversity of research designs and the diversity of the quality of research methodologies. However, according to Faria (2001), it seemed knowledge, ability, motivation, and focused efforts help simulation players to perform.

Because it seemed reasonable that simulation performance was a proxy for learning and because an extremely high percentage of instructors who use total enterprise simulations in college capstone courses grade on simulation performance (Anderson & Lawton, 1992b), the relationship between learning and performance has been relatively heavily researched. If research studies determined there was a strong correlation between the two, scholars could use performance increases as a way to assess the efficacy of the teaching approach—in effect, claiming high performers also learned the most. However, research on this relationship has strongly suggested the two variables did not covary (Anderson & Lawton, 1992a; Thorngate & Carroll, 1987; Washbush & Gosen, 2001; Wellington & Faria, 1992). Thus, performance is not a proxy for learning, and it is inappropriate to assess simulations using performance as a measure of learning. This is unfortunate for the management education and development field because for simulations at least, a simulated firm's economic performance in the form of such variables as market share, profit margins, forecasting accuracy, and stock price is accessible and easy to measure whereas learning is not. As a result, the field has turned away from raw or simple company or team performance and more toward measuring teaching approach effectiveness.

## The critiques and prescriptions of simulation validity

For much of their existence, the simulation approach has had to weather storms of criticism, especially criticism concerning their validity. Burns, Gentry, et al. (1990) decried the lack of rigorous research in evaluating the learning effectiveness of experiential methods. The focus of their concerns was exactly the issue of validity: internal validity or changes in the student in classroom settings, external validity or the generalizability of any learning effects to outside classroom situations, and transfer-internalization validity or the ultimate impact of the simulation experience on the student's career.

Carvalho (1991) echoed the flavor of these criticisms by noting there was no generally accepted theory and methodology for validating business simulations. More recently, Wolfe and Crookall (1998) continued this discussion by blaming problems on a lack of a generally accepted typology that captures the essence of the gaming experience; the simulation field's eclectic, interdisciplinary nature; and skepticism about the impact of educational research on good teaching practices. Feinstein and Cannon (2001, 2002) also contributed to this literature by tracing the problem of evaluating simulations on the existence of a cluttered literature that demonstrates a lack of consistency in defining terms and concepts.

Along with criticizing the field for not achieving successful validity research, many authors have offered prescriptions for creating valid research. Cannon and Burns (1999) proposed a process for validating gaming experience linked to actual demands of managerial experience, including performance measurement. This argument was also reflective of all elements of the Bloom taxonomy (Bloom et al., 1956). Feinstein and Cannon (2001, 2002) argued simulations should exhibit fidelity, the experience's level of realism, verifiability, having the game's model operate as intended, and validity, or having the conclusions reached by participants being similar to those that would be obtained from a similar real-world situation. They proposed three systems are at work—development, validation, and educational—and that these three systems should be related through two forms of validation. The following relationships were defined:

1. The relationship between development and validation used the term *representational validation*, which is proving the simulation does what it intends to do.
2. The relationship between validation and the educational method itself is represented with the term *educational validation*, which is proving what is being learned is consistent and relevant.

According to these authors, rigorous attention to these details would provide a basis for the proper evaluation and measurement of simulation learning.

## The studies

Although one would not know it from the aforementioned criticisms, there have been a reasonable number of studies whose purpose was to determine the validity of simulations. There have been five types.

*Studies assessing the validity of games in strategic management/business policy courses*. Because simulations are often used in business capstone strategic management courses and because total enterprise simulations are strategic in nature, the complementary nature of simulations to course content is a logical concern. There has been a series of studies using objectively measured, course-related learning outcomes in controlled or equivalent group research designs whose purpose was to assess game validity in strategic management courses. Wolfe (1997) reviewed these studies and concluded simulations assisted students in learning strategic management and were therefore valid. These studies suggested simulations have (a) representational validity (Feinstein & Cannon, 2002) in that they indicated the simulation did what was intended and (b) validity (Feinstein & Cannon, 2002) in that conclusions reached by participants are similar to those obtained from the real world being modeled.

*Studies assessing the simulation in terms of skills the simulation author proposed would be accomplished*. Carvalho (1991) concluded THE EXECUTIVE GAME (Henshaw & Jackson, 1984) was valid because it was stable in terms of the exponential forecasting skills and elementary skills associated with planning and control. Washbush and Gosen (2001) focused on learning increases resulting from a simulation itself, which were the learning objectives stated by the authors of MICROMATIC (Scott, Strickland, Hofmeister, & Thompson, 1992). They found before/after learning scores increased significantly for 9 of 10 data sets and therefore argued learning took place and that the simulation was a valid learning methodology. Feinstein (2001) focused on dynamic learning increases resulting from playing a food service operation simulation called FIST. This simulation consisted of a case study that introduced the simulation situation followed by simulation play (Feinstein, 2001). He found students in the FIST experience showed greater learning from before to after the experience than did those in a control group. These studies imply the simulations used had representational validity in that they indicated simulations did what was intended.

*Studies focusing on the performance of players who actually make decisions as opposed to making random decisions*. Dickenson and Faria (1997) examined the performance of "real" simulation companies versus the performance of a group of companies using random strategies. The real companies outperformed the random strategy companies, and thus they argued, the simulation used was internally valid. Wolfe and Luethge (2003) performed a similar study. They found real players who consciously made decisions outperformed both players who copied the industry leader's decisions and players who simply regurgitated their opening decisions. These studies appeared to indicate the simulations exhibited fidelity or an appropriate level of realism.

*Studies comparing player decisions with those expected by the simulation*. Dickenson, Whitely, and Faria (1990) detected students facing a "pull" environment in the marketing simulation LAPTOP (Faria & Dickenson, 1987) were not making significantly different decisions than those facing a "push" environment, which implied the simulation was not valid because player behavior was not appropriate to the situation faced. Wellington and Faria (2001) found introductory marketing students playing a simple simulation did perceive their environments were changing but could not identify the nature of the changes. This also implies the simulation was not valid although a degree of sense making was occurring. In these two cases, the games lacked representational validity because the simulations did not bring about their intended results.

*Studies assessing external validity.* Wolfe and Roberts (1986, 1993) performed two studies comparing the economic performance of teams and individuals with that of career-related results 5 years later. In the 1986 study, they found positive correlations between simulation performance and workplace promotion rates and career satisfaction. In the 1993 study, within-team game-based peer assessments were compared to later career mobility, career satisfaction, and salary levels. Again they found significant relationships between simulation players' influence on decision making, leadership, and esteem and value to the team on one hand and the persons' later income levels and salary increases on the other.

There is another set of studies that reflect on the external validity of simulations. These have shown the personality characteristics and decision-making styles used by successful real-world executives are similar to those used by successful simulation players and groups in total enterprise management games. Babb, Leslie, and Van Slyke (1966); Gray (1972); McKenney and Dill (1966); and Vance and Gray (1967) found the traits of successful game players were similar to those of successful executives. Wolfe (1976) found the simulation-playing behaviors of successful student players were the same as those of practicing middle managers. The implications of these studies are that simulations were externally valid, or in Feinstein and Cannon's

(2001) terms, possessed educational validation by indicating the learning derived from playing them is consistent and relevant.

In summary, we note there is evidence suggesting simulations are valid. However, although the Dickenson and Faria study was creative and although the studies Wolfe reviewed used controlled or equivalent group research designs, there does not exist enough consistent research from methodologically sound studies across multiple games to conclude simulations are valid. One could argue there are enough methodologically sound studies to tentatively conclude simulations are a valid method to teach strategic management. However, for studies assessing the simulation in terms of the skills simulation authors have proposed for them, a conclusion indicating games are valid would be unwise as there are only three such studies and each assesses learning from a different simulation. Furthermore, there are only two studies indicating that real, conscientious players outperform random, less conscientious players and only two that compare the behavior of players with the same individual's level of real-world success. Candidly, these positive and supportive findings are not enough to claim general validity for all simulations. This is not to say simulations are not valid, but instead, there is not enough evidence to conclude they actually are valid.

## Measurement issues

Beginning with this section, we will combine our observations regarding simulations and experiential exercises. Up to this point we have focused on the field's computer-based simulation research because this research is rather focused, whereas the field of human-based experiential learning research is just the opposite. Simulations, especially business simulations, refer to a restricted set of phenomena. To simulation scholars, it is clear what a simulation is and what it is not. Simulation researchers do almost all their work in business courses, and therefore a definite, practical set of outcomes is supposed to occur. Also, there is a performance component to most simulations, at least those developed in North America. Simulation researchers talk to each other and build off of each other's work, and the real-world and practical use of business games is one of the issues simulation scholars discuss to a great degree.

Conversely, experiential exercises are used in a vast and diverse number of environments. A tremendous number of them could be classified as experiential learning environments from accounting practice sets to mountain climbing. Outcomes from these experiences are concomitantly diverse, and there is only occasionally a "hard" performance component included in their design. In some disciplines, experiential exercise scholars build off each other's work, but this is not done to the same degree as with computer-based simulation scholars. Furthermore, it appears it is not done at all across academic disciplines.

So for computer-based simulations, there is a concentrated body of scholarly concern in which performance, validity, and cause and effect are discussed. And although experiential scholars are interested in assessing their teaching approach, a concentrated group with common conceptual concerns does not exist, and validity is not an explicit concern among them. Therefore, validity discussions exist in the field of simulation research and not in the field of experiential learning research, and we are able to present scholarship on the validity of simulations and not on the validity of experiential learning. Alternatively, the same measurement issues arise in both computer- and human-based experiential milieus, thus we are able to present scholarship on measurement issues for both kinds of experiential learning.

### The critiques and prescriptions

Scholars concerned about the measures and instruments used to assess experiential learning and simulations agree to a great extent as to what is wrong with past assessment attempts and what should be done to correct the situation. We note four generally accepted prescriptions. First, research design issues should be attended to with pretests and posttests, treatment and control groups, experimenter control of treatment variables, and random assignments to groups (Campbell & Stanley, 1963). Second, outcome variables need to be defined, objective, and appropriate for the experience being assessed. J. W. Gentry et al. (1998) along with B. Keys and Wolfe (1990) suggested the problem with assessment research is rooted in an incomplete definition of the nature of learning and a lack of systematic efforts to obtain these objective measurements. Burns, Gentry, et al. (1990); Butler et al. (1985); and Anderson and Lawton (1997) stated the best way to assess these approaches is by defining what is to be learned from a teaching objectives standpoint and by developing objective measures of the construct to detect if the participants learned what they were supposed to from the experience. Some scholars have stated that many outcomes that seemingly measure learning are inadequate. For example, Wolfe (1990) argued grades and tests unconnected to learning goals are unproven proxies for learning, and J. W. Gentry et al. (1998) pointed out learner perceptions of what they state they have learned are inadequate and invalid.

J. W. Gentry et al. (1998) included the too commonly used "perception of cognitive learning" type measure as a "feel good" measurement category. If the students enjoyed the experience, there is a likely halo effect in terms of measuring perceived learning. J. W. Gentry et al. also asserted perceived learning and objective learning indicators measure separate constructs. The findings by Gosen and Washbush (1999) found support for their assertion. In their study, a zero correlation existed between objective learning scores and the perceived learning scores on 10 types of learning measured.

The third prescription regarding measurement is that learning measures should be tied to explicit learning goals. According to Anderson and Lawton (1997),

> Much of the reason for the inability to make supportable claims about the efficacy of simulations can be traced to the selection of dependent variables and the lack of rigor with which investigations have been conducted. . . . Virtually all research designed to measure the outcomes produced by engaging in an activity requires . . . assumptions concerning the expected outcomes produced by performing that activity. We cannot construct an assessment activity without knowing what it is we expect to measure. (p. 68)

Finally, the measures should be valid. Instrument validity is a priority for measurement throughout the social sciences. The test score's validity, according to McDonald (1999), is the extent it measures an attribute of the respondents that the test is employed to measure for its intended population. Simply stated, a test is valid if it measures what it purports to measure. Any instrument should present proof that it measures what it is supposed to measure. Anderson, Cannon, Malik, and Thavikulwat (1998) proposed the following evidential validity assessment instrument standards:

1. Show evidence of reliability between the results obtained at one time to those obtained later when applied to the same subject.
2. Be able to discriminate between individuals possessing different skills or performance levels.
3. Show convergence with other instruments measuring the same constructs.
4. Yield normative scores for different populations.

## The studies—Experiential exercises

This section focuses on the merit of the variety of devices used to assess the effectiveness of the field's experiential exercises. More than 80 studies were read for this review of experiential studies, and 19 of those are cited in this article. All studies were from published journals, and most came from nonbusiness fields. All were written since 1989. Those prior to 1989 have been included in the assessment of experiential learning article in the *Guide to Business Gaming and Experiential Learning* (Gosenpud, 1990). No studies were included from the ABSEL proceedings because none assessed experiential exercises in terms of (a) objective learning measures, (b) an observable behavior, and/or (c) the use of a pre-post or control group design. Studies included in this article met the three aforementioned criteria while also having been published in a journal since 1989.

The studies assessing experiential pedagogies have been undertaken in a wide variety of learning environments. The ones found for this article have focused on experiential learning with among others, medical, dental, and nursing students; with a statistics class; in outdoor training programs; with service learning among ninth graders; for problem adolescents; in an online professional development course; for social work students; for undergraduates in a leisure and recreation course; and in an environmentally oriented work program. The search of studies for this article has been extensive and broad but not exhaustive. Thus, it is possible that a number of studies assessing experiential pedagogy in other learning milieus may have been overlooked.

The experiential learning courses and programs that were found have been assessed in a variety of ways. Blake (1990); Burns, Rubin, and Bojack (1990); Specht and Sandlin (1991); Premi and Shannon (2001); and M. Gentry, Iceton, and Milne (2001) used objective learning tests. Blake determined students in an experiential organizational behavior section had higher test scores than students from lecture sections, and M. Gentry et al. showed higher pre/post knowledge levels of the nature of community service after a 3-day course. Burns, Rubin, et al. (1990) used an objective test based on the Bloom cognitive domain taxonomy and found those participating in a computer-aided experiential exercise attained higher test scores than those participating in a non-computer-aided exercise. Specht and Sandlin similarly used objective tests to examine knowledge acquisition in an accounting class 1 and 6 weeks after an experiential intervention. Test scores for those receiving the intervention did not decrease during the 6 weeks between tests, but the lecture-taught control group's test scores decreased significantly. Premi and Shannon studied the effects of a continuing medical education video program for practicing physicians. Assignments to program and control groups were randomized, and A/B forms were used as crossed pretests and posttests for half of each group. The test, which covered five medical knowledge domains, was tested for reliability and face and content validity by practicing physicians (Premi & Shannon, 1993). Significantly improved test performance was associated with those participating in the program with no improvement for the control group.

Dedeke (1999), Rocha (2000), James (2000), and Beaumie, Williams, and Dattilo (2002) also assessed learning but used perceived learning measures. All participants reported learning gains. In Dedeke, students said they learned from doing a response paper. Rocha's learners perceived themselves to be better communicators and media users. James's learners stated they had a deeper understanding of environment-oriented work, and in the Beaumie et al. study, leisure and recreation students believed they could apply concepts in real-world situations.

Some courses and programs were assessed using behavioral measures. Herbert and Landin (1996) measured tennis performance after manipulating practice schedules. M. Gentry et al. (2001) found their students improved in five skill categories, including feedback and reassurance, as evidenced by structured role plays. Dedeke (1999) found participation in a journal-writing exercise produced shorter papers in a management class than nonjournal writers. Rocha (2000) detected students taking an experiential service learning policy class in a social work school were significantly more likely to be members and be active in coalitions or committees than graduates of the same school who did not take the service learning class.

Experiential learning courses and programs have been assessed in a variety of other ways. Experiential course programs or modules have reported increased confidence (Manoque, Brown, Nattress, & Fox, 1999), enjoyment (Dedeke, 1999), and moral reasoning (Smith, Strand, & Bunting, 2002). As a result of experiential learning practices, other studies have reported perceived increases in skills as defined by the Kolb (Boyatzis & Kolb, 1991) experiential cycle (Brodbeck & Greitemeyer, 2000; Herz & Merz, 1998), perceived increases in group cohesion (Glass & Benshoff, 2002), and increases in self-regulation and peer self-esteem for alternative-learning high school students (Nichols & Steffy, 1999).

## Computer-based simulations

Although simulations have also been used in a variety of learning milieus including pilot and military training programs; sociology, meteorology, and ecology courses;

and business classes, simulation assessment attempts have generally been undertaken within a narrower range of environments. Our search turned up studies only in the fields of business and food preparation. About 35 simulation papers were read, and 20 were included in this article. Our search criteria for computer-based simulation studies differed somewhat from those for experiential exercise studies. First, older studies were included because end results from those studies integrated and added considerably to the points we wanted to make for this section. Second, studies from the ABSEL conference proceedings were included because some of them used comparison group, control group, or pre-post designs. As with experiential assessment articles, we included only those with control groups or pre-post designs whose research design assessed a simulation experience and yielded quantitative results.

Similar to experiential exercise assessment studies, computer-based simulation assessment studies have used a variety of measures to assess their effectiveness. Raia (1966) used course exams and found game-playing groups outperformed controls on a posttest consisting of true-false items, a break-even analysis, and a sales forecasting exercise. Faria and his colleagues (Faria & Whiteley, 1990; Wellington & Faria, 1991; Whiteley & Faria, 1989), Wolfe and Guth (1975), Wolfe (1976), and Wheatley, Hornaday, and Hunt (1988) also used course exams, but these authors differentiated exams into general categories of objectives prior to test construction. For Faria and colleagues, the categories were quantitative, applied, and theoretical items. Whiteley and Faria (1989) found students playing a simulation scored significantly higher on quantitative questions than students not playing a simulation but no higher on applied questions. A year later, Faria and Whiteley (1990) determined students playing a simulation scored higher on quantitative and theoretical final examination items but not higher on applied items. One year later, Wellington and Faria (1991) found results that contradicted the earlier two studies. In this latter study, simulation students did not perform better than a control group on an examination when grade point average was factored into the results.

For Wolfe and Guth (1975), the categories were policy-making principles and policy-related facts. These authors gave essay exams that were blind graded to two sections of a strategic management course. One used only cases, the other a simulation. Students in the simulation section outperformed case students on policy-making principles but not on policy-related facts. Wolfe's (1976) focus was on the effects of game participation on learning strategic management and organizational goal setting. His objectives included administrative skills, environmental cognizance, and policy system knowledge. He found students and middle-level managers scored significantly better on posttests measuring administrative skills and environmental cognizance than on pretests. Wheatley et al. (1988) focused on goal-setting ability as an objective, and students exposed to a simulation learned more than students exposed to cases.

Whereas the aforementioned studies have focused on a limited set of learning objectives, two studies assessed simulation learning in terms of a greater number of objectives. In addition, in both, instruments were developed on which validity studies could be performed. Feinstein (2001) developed an examination instrument testing whether or not students learned from the FIST simulation. The items were intended to reflect mastery according to Bloom's cognitive domain taxonomy (Bloom et al., 1956) as well as the simulation's inherent goals. In this study, 20 food service graduate students validated the instrument by judging whether the content was consistent with the simulation, and 20 instructional system PhD students judged whether each question fell within the ascribed classification according to Bloom's taxonomy. Feinstein found the FIST experience resulted in increased learning. Those students who participated in FIST scored significantly higher on a posttest than they did on a pretest, whereas a control group showed no differences between pretests and posttests.

Gosen, Washbush, and colleagues (Gosen et al., 1999; Gosen & Washbush, 2002; Gosen, Washbush, & Scott, 2000) performed a set of studies whose purpose was to develop a valid instrument for assessing learning associated with the MICROMATIC simulation (Scott et al., 1992). These authors developed a set of 40 learning objectives derived from simulation experts and the simulation's authors (Scott et al., 1992). From those objectives, a comprehensive set of 122 examination questions (Gosen et al., 1999) was developed. This effort led to validity studies of five alternative examination forms using many of the original 122 questions (Gosen et al., 2000; Gosen & Washbush, 2002). The studies tested two of the form's convergent validity with forecasting accuracy and self-reports of learning as criterion variables. The results were negative because the learning scores and criterion variables were not positively correlated.

There was one study that employed proxies of learning to assess simulations. Peach and Platt (2000) used forecasting accuracy and performance improvements as proxies for learning and found forecasting accuracy improved substantially in periods when performance was staying static and vice versa. From this they concluded that the early stages of simulation play set up learning in later stages.

Some studies used perceptions of learning to assess simulation effectiveness. McHaney, White, and Heilman (2002) differentiated successful and unsuccessful simulations according to student perceptions. Herz and Merz (1998) found students participating in a simulation believed their learning experience contributed to all stages of Kolb's learning cycle (Boyatzis & Kolb, 1991) to a greater degree than students who participated in a traditional seminar. Comer and Nichols (1996) found simulations to be a helpful aid to both content and process learning in the learners' eyes.

Others have focused on the attitudes toward simulations. Tompson and Dass (2000) found simulation participation resulted in significantly greater self-efficacy. Leonard and Leonard (1995) determined business school graduates believed courses using simulations did a better job of preparing them for their current job than courses using only cases. Hergert and Hergert (1990) assessed simulation effectiveness in terms of perceptions of course structure, game parameters, and effort and performance exerted. Washbush and Gosenpud (1991) and White and Von Riesen (1992) found satisfaction increased with simulation participation, and Zalatan and Mayer (1999) found improvements in end-of-year surveys after simulations were included in the curriculum.

**TABLE 1: Experiential Exercise Assessment Studies Reviewed**

| | *Assessment Focus* | | |
|---|---|---|---|
| *Research Design/Measurement Category* | *Learning* (N = *9)* | *Behavior* (N = *4*[a]*)* | *Other* N = *7*[a]*)* |
| Objective measure(s) | 5 | 2 | 0 |
| Self-report | 4 | 0 | 7 |
| Judgment | 1 | 0 | 0 |
| Proxies | 0 | 0 | 0 |
| Pre/post measure(s) | 1 | 0 | 0 |
| Control or comparison groups | 3 | 0 | 0 |
| Pre/post plus control or comparison groups | 1 | 0 | 0 |
| Objectives-driven instrument | 1 | 0 | 0 |
| Instrument validation | 1 | 0 | 0 |
| Convergent validity study | 0 | 0 | 0 |

a. One study assessed both behavior and nonlearning phenomena.

## A summary of the research

At face value, the studies just reviewed showed clear support for the effectiveness of computer-based simulations and experiential exercises. All the experiential exercise and all but two of the computer-based simulation studies revealed that learning took place, although in some of the computer-based simulation studies, learning occurred in certain categories and not others. In addition, learners exposed to both versions of the experiential method reported they learned and that their experience was positive. These learners also reported positive attitude changes. As a result of experiential learning, changes in behavior and intrapersonal effectiveness were also reported.

Although all this appears to be a glowing endorsement of computer-based simulations and experiential exercises, an examination of those studies using an adherence to rigorous measurement and research design standards yields a more complex set of conclusions. Tables 1 and 2 summarize the assessment studies reviewed here using Anderson et al.'s (1998) criteria for studies using measurement instruments to assess learning approaches. Table 1 summarizes how the experiential exercise studies fared, and Table 2 summarizes the results for the computer-based simulation studies. There were 19 experiential studies covered in this article. The columns are divided according to dependent variable. The rows reflect scholars' prescriptions regarding learning approach assessment using measuring devices.

For the experiential-exercise-associated learning in Table 1, 9 studies focused on learning. Five studies used objective learning measures, and 4 used self-reported learning measures. Four assessed experiential learning in terms of behavior, and 7 assessed in terms of perceptions of nonlearning phenomena. Of the total number of experiential exercise studies reviewed, 11 used self-reports, 4 used comparison groups

**TABLE 2: Computer-Based Simulation Assessment Studies Reviewed**

| | *Assessment Focus* | |
|---|---|---|
| *Research Design/Measurement Category* | *Learning* (N = *14)* | *Attitudes* (N = *6)* |
| Objective measure(s) | 10 | 6 |
| Self-report | 3 | 0 |
| Proxies | 1 | 0 |
| Judgment | 0 | 0 |
| Pre/post measure(s) | 2 | 0 |
| Control or comparison groups | 6 | 1 |
| Pre/post plus control or comparison groups | 2 | 0 |
| Repeated measures | 1 | 0 |
| Objectives-driven instrument | 2 | 0 |
| Instrument validation | 2 | 0 |
| Convergent validity study | 1 | 0 |

without pre-post testing, and 1 used instruments derived from predetermined learning objectives.

Of the 20 computer-based simulation studies presented in Table 2, 10 used objective learning measures, 3 employed self-report measures, 1 used learning proxies, and 6 used player attitudes as the assessment basis.

Although these tables show some totals, they do not reveal the true picture of the quality of any individual study. For example, there were nine papers assessing experiential exercises in terms of learning, and of the nine, five used objective measures and four self-reports. Of these same nine studies, three used a comparison method, one used a pre-post method, and one used a pre-post comparison group method. Of the nine, only one used an instrument developed from predetermined objectives, and of the nine, only one developed validity studies for the instrument employed. Thus, any given study using an instrument developed from predetermined objectives may or may not have used a pre-post design and may or may not have performed a validity study on the instrument employed.

For our review, we read more than 115 articles and chose 39 to be included in this article. These were chosen because they more closely met the highest research standards needed if validity claims are to be supported. Tables 1 and 2 show that of those 39, 9 used a pre-post design, 12 used a comparison group design, and 4 used a combination pre-post, comparison group design. Seventeen used objective measures, 9 used instruments derived from predetermined objectives, and in 3, validity studies were performed. What is more important and not reflected in Tables 1 and 2 is that none of the 115 studies considered for this article met all the criteria for sound research.

What the tables show is that in only 3 of the 39 studies reviewed were validity studies performed. Thus, it can be argued none of the others assessed the teaching approach properly. Of the 3 in which validity studies were performed, the Gosen/Washbush series never used a comparison group, and the studies using criterion variables failed to show the learning measure was valid. Both Feinstein's and Premi and Shannon's validity studies can be considered of the face-validity type, which is arguably defective. Thus, using the relatively rigorous standards that stem from the Anderson et al. (1998) prescriptions considerably weakens how firm our conclusions can be about the teaching effectiveness of the two approaches studied. We probably can say there is evidence the approaches are effective, but the studies showing these results do not meet the highest of research design and measurement standards. Thus, we believe any conclusion about them must be tentative.

## Discussion and conclusions

Our purpose is not to find flaws with studies reviewed. For example, both Premi and Shannon's and Feinstein's studies were carefully conducted and thoroughly done, the results of which rightfully contribute to our understanding of the value of experiential exercises and computer-based games. In addition, it could be argued the studies reviewed here produced tentative conclusions supporting the value of these versions of the experiential approach. Our point is that little has been actually done to design studies and develop instruments that could competently assess teaching approaches despite a common understanding of what should be done and all the criticisms and prescriptions expressed by scholars in this field and others.

Given our assessment of the situation, it is important to understand why the highest research standards have been relatively rare—and determining the why might be a useful research avenue itself. We propose three reasons. First and frankly least important, careful, rigorous research dedicated to developing a valid instrument and reflective of thought-out learning objectives is extremely time-consuming. Perhaps too time-consuming to be worth the effort. For us in our attempt to conduct a validity study, the process meant asking experts for objectives and test items. It also meant testing an array of test items on student samples to determine clarity, difficulty, and covariance among items. It meant asking colleagues to participate in a number of interuniversity studies. It also meant doing pilot studies and refining criterion variables. Sad to say, it is far, far easier to read Campbell and Stanley (1963), offer prescriptions, and then tell the world's researchers they have failed to follow these prescriptions than to scratch through the process of actually doing the research.

Second, the criterion variable being used, which is learning from a computer-based simulation or experiential exercise, is illusive. To our knowledge, every attempt to concretize this variable has failed. As a field, we believe we know what it is, but what it looks like so it can be measured lacks form. Recall that one attempt to measure learning via the proxies of performance and forecasting accuracy did not correlate with learning scores. The problem lies at least in part in the nature of learning and the learners' expressions of it. Learning is an internal mental process, and what is learned and how it is learned is unique for each individual. To create an instrument able to capture what is actually learned is easier said than done. To get at this, learners have to be motivated to express their learning, and an instrument applied at the experience's end, even for a grade, may be neither a proper vehicle for that expression nor sufficiently motivating.

Third, our present world does not appear to care enough for this work to be successful, and perhaps it should not. Apparently, now is not the only time scholars believed this. In quoting educational scholars of the 1930s, Wolfe and Crookall (1998) revealed educational research at that time was judged to be petty, tedious, equivocal, unreplicable, and inconsequential. Some would say the situation has not changed that much over the years as the number of scholars engaged in empirical validity research has rarely been large. As a personal example, in the late 1990s we asked colleagues to help us test the instrument reported in Gosen et al. (1999, 2000) and Gosen and Washbush (2002) over a 3-year period. We obtained four initial volunteers, but only one carried through on the informally agreed-on participation contract.

Another example of a failed ABSEL-related interuniversity research initiative is that associated with the ABSEL Research Consortium (Wolfe, 1986). Although starting with a dedicated conference session during which 11 founding position papers were presented, the project was abandoned after 2 years of inactivity (Burns, 1986; Cooke, 1986; Frazer, 1986; Fritzsche, 1986; Gold & Pray, 1986; Goosen, 1986; Gosenpud, 1986; Graf & Kellogg, 1986; J. B. Keys, 1986; Kline, 1986; Norris, 1986). It appears there is just too much work involved and too little a pay-off for most business educators. In all honesty, despite the moral invectives that could be raised, it appears few care if the instruments being used to measure learning effectiveness are valid. Who will lose tenure or will be thought of as being a bad person if the instruments used are not validated by formal validity studies? Most of us who use the experiential method believe it is an excellent teaching vehicle and do not need rigorous research to substantiate or bring comfort in that belief. Both computer-based simulations and experiential exercises are established enough so assessment research that verifies their credibility is unnecessary. Furthermore, very few have called for validity studies to prove the value of lectures, cases, quizzes, examinations, experiments, discussions, book reports, term papers, or any of the other typical class methods being touted by their advocates. Why in the world should we hold experiential learning or simulations to higher standards?

Before we leave this discussion of why there have been so few studies using accepted research design standards, it should be noted there is a parallel field in which empirical studies have been undertaken to determine the effectiveness of training/teaching efforts. This field is managerial training. Like experiential learning, it is very diverse and one in which there is an effort to train or teach concepts, attitudes, perspectives, behaviors, and/or skills. Many managerial training programs use experiential techniques, but there are differences between the two fields. On one hand, experiential learning is primarily an academic field. Professors undertake the research, and their criterion variables usually include learning. On the other hand, managerial training takes place in businesses. Studies are performed by practitioners as well as academics, but learning is less likely the primary criterion variable. So although there are similarities between the two fields and some overlap, there are also differences.

In 1986, Burke and Day performed a meta-analysis of 70 studies measuring the effectiveness of six types of management training. These covered such training methods as role-playing, lecture, discussion, wilderness exercises, and sensitivity training, covering such skill areas as human relations, leadership, problem solving, and decision making, with the programs using any combination of four types of measurement techniques to determine if the program was successful. All studies had as their subject managerial or supervisory personnel. All used comparison groups and measured effectiveness with quantitative measures. Of these studies, 18 used objective learning criteria and 8 used objective behavioral criteria. The results demonstrated training effectiveness to a reasonable extent when subjective learning and objective behavioral criteria were used and to a small extent when using objective learning or subjective behavioral criteria.

We in the experiential learning field can learn from Burke and Day (1986). These authors studied a diverse field and categorized the field's research so it could be viewed holistically. This also can be done for experiential learning. Their review supports the notion that accepted research standards can be used to assess the effectiveness of experiential programs. Our fields should be encouraged by the existence of this review.

To begin summarizing our thoughts, after reviewing a considerable amount of recent literature assessing computer-based simulations and experiential exercises, we have come to a disappointing conclusion. Based on Bloom's taxonomy (Bloom et al., 1956; Krathwohl et al., 1964) and rigorous research design standards, there have not been enough high-quality studies to allow us to conclude players learn by participating in simulations or experiential exercises. Given the same research design standards and despite a considerable amount of affective research, there remain open questions as to how participants feel about simulation participation and whether or not what they think they are learning is correlated with actual learning. Similarly questionable are attempts to measure or draw conclusions about postplay behavior. The general findings of such research streams are that simulations and experiential activities are generally accepted, there is a mild preference for simulations over other experiential modes, and there are positive learning effects—and we cannot even say this for sure because there are too few studies that used comprehensive research designs.

With respect to the motivation to learn, nothing has been proposed that offers anything new. The postulates of arousal theory (Berlyne, 1967) with respect to the attractiveness of challenge and the concept of the learning space (goal, valence, psychological distance), as proposed long ago by Lewin (1935, 1936), remain relevant. However, we might ask, do we really need to be reminded that if students believe that trying to do something that is interesting and important they will both make an effort and believe it was worthwhile? It remains doubtful that operationalizing a concept of learning motivation is possible in any global way. From a practical standpoint, this is a problem all teachers face all the time, and most of them are probably able to deal with this reality in practical ways.

All these topics are important. However, from a teaching/learning perspective, they tell us little. Can any educational researcher ever trust anyone who in a burst of enthusiasm after taking a course or completing an exercise exclaims, “I learned so much” but after this glowing generality cannot describe anything of substance about what was learned and how it was absorbed?

Learning does seem to be the important issue for educators, but it is only recently beginning to receive our field’s systematic attention. Validation of learning is a current major focus of concern. Unfortunately, with respect to this topic, it is far too easy to criticize, intellectualize about educational purity, and propose increasingly cumbersome theoretical models than to undertake the necessary work to prove or disprove a particular learning approach or strategy. Much of the research entailed in validating learning will come, of necessity, from classroom settings, and given the paucity of studies up until now and the inconclusiveness of the results, a difficult challenge awaits the learning assessment field. It is a major hurdle.

Why are the results of the studies we have reviewed so lacking in consistency and rigor? To suggest an answer, we turn to a contemporary critic, E. D. Hirsch, Jr. (2002), who believes educational research has followed the form but not the essence of science. He stated, “Educational data are difficult to apply in a dependable way because of the contextual variables that change from classroom to classroom and year to year, and that drown out the effects of single or multiple interventions” (p. 4). This is precisely the dilemma confronting those who desire to validate and measure experiential learning methods. Typically, studies are performed by different scholars, on different campuses, with different types and groups of students, in different courses, all using a wide variety of simulations and exercises. Given this state of affairs, is it even worthwhile to pursue the rigorous validation of the learning potential of simulations, games, and exercises?

One approach would be to simply avoid the issue and to assert as an act of faith that beyond a narrow range of likelihood, using simulations, games, and exercises will inevitably induce learning that is not defined by instructional objectives and design features and that may not even be directly or easily measurable. We in fact argue that educators ought to allow for that kind of learning as opposed to trying to define everything with precision.

There is unfortunately little evidence, with a precious few exceptions, that anyone is serious about jumping through the necessary hoops, most of which are filled with semantic landmines, to change the current situation. Most of those who propose scientific purity are not likely to go beyond their own arguments or take the steps needed to implement the processes they say are important and meaningful. What we are faced with is little more than the educational equivalent of an old-fashioned fire-and-brimstone sermon in which a preacher tells us what we have done wrong, how things should be, and now it is up to us to effect the changes. One would like to see the purists practice a little purity on themselves.

Does this mean we are becoming educational Luddites and are simply throwing our hands up and saying, “Forget it, it’s too hard”? Perhaps, but only perhaps. Although we are not likely to transcend all the limitations identified by us and Hirsch (2002), a number of actions are possible and necessary. First, there needs to be clarity provided by simulation and exercise developers concerning what the exercise attempts to provide and how. Experiential users should have access to sufficient information to intelligently and logically be able to select simulations and exercises for student use—some ability to decide whether or not a given method is usable in a specific learning situation. Beyond that essential, basic point, those who do research relating to any form of experiential learning need to begin to collaborate in comprehensive ways to provide ways to measure outcomes and have the potential to be validated. Wolfe and Rogé (1997) described for us a variety of games that more or less relate to a spectrum of educational purposes, and therefore users could select from them accordingly. Whether to go beyond that and pin down validation is a further question for the field.

There is ample evidence at least for simulations that performance and learning do not covary. The importance of this cannot be overstated. That is exactly what we should find—that the simulation is in fact a learning experience for all participants no matter how well any given person performs. We need the courage and determination to go forward. What are needed are broad-based, large-scale projects involving many scholars from a wide variety of disciplines and educational settings designed as carefully as possible to move us forward. To that end, ABSEL's Legacy Project and the Bernie Keys Library, two new efforts described in this symposium issue by Cannon and Smith (2004), are steps in the right direction. This effort has not yet produced any new validity studies, but the promise is there.

The problems identified by Hirsch (2002) will not easily be transcended, but a determined effort can and should be made to probe learning potential and demonstrate that it has occurred. We argue we are not a surrogate Educational Testing Service attempting to measure performance and ability on easily definable cognitive areas, and we are not able to define things in a yes/no, either/or way. But as a field, we should be willing to take on the challenge. We hope there exists a considerable number who are concerned with building a reasonably broad base of relevant knowledge coupled with the ability to use it in constructive, creative, and effective ways in both school and postschool settings. From that perspective, learning environments need to be both bounded and unbounded. The instructor has to make the final decisions on where the lines of demarcation lie.

To take an optimistic perspective, undertaking research to determine the effectiveness of our approaches is very doable. We have the talent and the standards, and there are journals willing to accept competent studies. It has been done. Although Premi and Shannon (2001) and Feinstein (2001) developed instruments validated with face-validity methods, the measures were carefully constructed and might be valid given construct validity studies. Wolfe and his colleagues (Wolfe, 1997) and Dickenson and Faria (1997) have also done work that holds promise for demonstrating validity for simulation learning measures. We believe those are good beginnings. We also believe cross-institutional interest would expedite success. If those who care would make the effort to provide research settings and collaborate on creating instruments, variable definitions, and research designs, we could be well on our way to determining our teaching method's validity. As we have said more than once in this article, it depends on whether people in the field care.

Less optimistically and more pragmatically, the most reasonable approach to using any experiential method is whether or not it appears to be efficacious in the environment designed by the instructor and at the same time provides a usable way to assess student learning. Even if we do not adhere to the strictest standards, users need reasonably objective ways to assess student participation in experiential learning environments. These assessments should be free from social desirability problems, relate to agreed-on learning objectives, and provide means for individual assessments of participants. Constructing large test banks among other things would be appropriate. Such tests should provide reasonably varied and validated items, with at least face validity, and be arranged topically or by generally agreed-to learning objectives. Such tools would permit users to assess game impact on each participant in relation to instructor-defined intentions. This would allow for ease of use in constructing parallel forms for use in pretesting and posttesting by instructors. A major fieldwide effort should be undertaken to collect data on test administration so we could develop profiles for specific simulations and exercises. If the data suggest a given simulation or exercise does what it purports to do, that would be a good enough beginning for now. We need to develop standards of reasonableness. However, despite some tentative, preliminary attempts to do this, there has been little expressed interest or involvement from the field.

All this is a big challenge. Will it be answered? It depends on the business gaming field. Our discipline's tradition is to write, theorize, criticize, and especially prescribe. Some people do empirical follow-up, but not many. The field is capable of change, but will it change? Our suspicion is, to quote the immortal words of Sonny and Cher, "the beat goes on"—and on and on . . . .

## References

Anderson, P. H., Cannon, H., Malik, D., & Thavikulwat, P. (1998). Games as instruments of assessment: A framework for evaluation. *Developments in Business Simulation & Experiential Exercises*, *25*, 31-37.

Anderson, P. H., & Lawton, L. (1992a). The relationship between financial performance and other measures of learning on a simulation exercise. *Simulation & Gaming*, *23*, 326-340.

Anderson, P. H., & Lawton, L. (1992b). A survey of methods used for evaluating student performance on business simulations. *Simulation & Gaming*, *23*, 490-498.

Anderson, P. H., & Lawton, L. (1997). Demonstrating the learning effectiveness of simulations: Where we are and where we need to go. *Developments in Business Simulation & Experiential Exercises*, *24*, 68-73.

Babb, E. M., Leslie, M. A., & Van Slyke, M. D. (1966). The potential of business gaming methods in research. *Journal of Business*, *39*, 465-472.

Beaumie, K., Williams, R., & Dattilo, J. (2002). Student perceptions of interactive learning modules. *Journal of Research on Technology in Education*, *34*, 453-473.

Berlyne, D. E. (1967). Arousal and reinforcement, In D. Levine (Ed.), *Nebraska Symposium on Motivation* (pp. 1-110). Lincoln: University of Nebraska Press.

Blake, C. G. (1990). The effects of instructional strategies on the learning of organizational behavior by a large university class. *Journal of Instructional Psychology*, *17*, 59-64.

Bloom, B. S., Englehart, M. D., First, E. D., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, Handbook I: The cognitive domain*. New York: David McKay.

Boyatzis, R. E., & Kolb, D. A. (1991). Assessing individuality in learning: The learning skills profile. *Educational Psychology*, *11*, 279-295.

Brodbeck, F. C., & Greitemeyer, T. (2000). Effects of individual versus mixed individual and group experience in rule induction on group member learning and group performance. *Journal of Experimental Social Psychology*, *36*, 621-648.

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology*, *71*, 232-245.

Burns, A. C. (1986). The creation and operation of a data-base system for ABSEL research. *Developments in Business Simulation & Experiential Exercises*, *13*, 102-105.

Burns, A. C., Gentry, J. W., & Wolfe, J. (1990). A cornucopia of considerations in evaluating the effectiveness of experiential pedagogies. In J. W. Gentry (Ed.), *A guide to business gaming and experiential learning* (pp. 253-278). East Brunswick, NJ/London: Nichols/GP.

Burns, A. C., Rubin, R. S., & Bojack, J. (1990). Computer-aided exercises vs. workbook exercises as learning facilitator in the principles of marketing course. *Developments in Business Simulation & Experiential Exercises*, *17*, 29-33.

Butler, R. J., Markulus, P. M., & Strang, D. R. (1985). Learning theory and research design: How has ABSEL fared? *Developments in Business Simulation & Experiential Learning*, *12*, 86-90.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook on research on teaching* (pp. 171-246). Chicago: Rand McNally.

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.

Cannon, H. M., & Smith, J. A. (2004). The ABSEL Classicos initiative. *Simulation & Gaming*, *35*, 294-307.

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

Comer, L. B., & Nichols, J. A. F. (1996). Simulation as an aid to learning: How does participation influence the process? *Developments in Business Simulation & Experiential Exercises*, *23*, 8-14.

Cooke, E. F. (1986). The dilemma in evaluating classroom innovations. *Developments in Business Simulation & Experiential Exercises*, *13*, 110-114.

Dedeke, A. (1999). Design, integration, and student evaluation of response papers in an introductory management course. *Journal for Education of Business*, *77*, 211-214.

Dickenson, J., & Faria, A. J. (1997). The random-strategy criterion for validity of simulation game participation. *Simulation & Gaming*, *28*, 263-276.

Dickenson, J. R., Whitely, T. R., & Faria, A. J. (1990). An empirical investigation of the internal validity of a marketing simulation game. *Developments in Business Simulation & Experiential Exercises*, *17*, 47-52.

Faria, A. J. (2001). The changing nature of business simulation/gaming research. *Simulation & Gaming*, *32*, 97-110.

Faria, A. J., & Dickenson, J. (1987). LAPTOP: A principles of marketing simulation. *Developments in Business Simulation & Experiential Exercises*, *14*, 53-56.

Faria, A. J., & Nulsen, R. O. (1978). A new research focus: An analysis of the simulation game user. *Exploring Experiential Learning: Simulations and Experiential Exercises*, *5*, 25-31.

Faria, A. J., & Whiteley, T. R. (1990). An empirical evaluation of the pedagogical value of playing a simulation game in a principles of marketing course. *Developments in Business Simulation & Experiential Exercises*, *17*, 53-57.

Feinstein, A. H. (2001). An assessment of the effectiveness of the simulation as an instructional system for food service. *Journal of Hospitality and Tourism Research*, *25*, 421-443.

Feinstein, A. H., & Cannon, H. M. (2001). Fidelity, verifiability, and validity of simulation: Constructs for evaluation. *Developments in Business Simulation & Experiential Exercises*, *28*, 57-62.

Feinstein, A. H., & Cannon, H. M. (2002). Constructs of simulation evaluation. *Simulation & Gaming*, *32*, 425-440.

Frazer, J. R. (1986). Simulation—Indoctrination or learning. *Developments in Business Simulation & Experiential Exercises*, *13*, 108-109.

Fritzsche, D. J. (1986). Funding for a research consortium on simulation and experiential learning. *Developments in Business Simulation & Experiential Exercises*, *13*, 106-107.

Gentry, J. W., Commuri, S. F., Burns, A. C., & Dickenson, J. R. (1998). The second component to experiential learning: A look back at how ABSEL has handled the conceptual and operational definitions of learning. *Developments in Business Simulation & Experiential Exercises*, *25*, 62-68.

Gentry, M., Iceton, J., & Milne, D. (2001). Managing challenging behavior in the community: Methods and results of interactive staff training. *Health & Social Care in the Community*, *9*, 143-150.

Glass, J. S., & Benshoff, J. M. (2002). Facilitating group cohesion among adolescents through challenge course experiences. *Journal of Experiential Education*, *25*, 268-277.

Gold, S. C., & Pray, T. F. (1986, March). *"Correct" algorithms for "correct" learning*. Paper presented at the Association for Business Simulation and Experiential Learning, Reno, NV.

Goosen, K. R. (1986). An analysis of ABSEL conference papers (1974-1985). *Developments in Business Simulation & Experiential Exercises*, *13*, 97-101.

Gosen, J., & Washbush, J. (1999). Perceptions of learning in TE simulations. *Developments in Business Simulation & Experiential Exercises*, *26*, 170-175.

Gosen, J., & Washbush, J. (2002). The validity investigation of a test assessing total enterprise simulation learning. *Developments in Business Simulation & Experiential Exercises*, *29*, 247-251.

Gosen, J., Washbush, J., Patz, A. L., Scott, T. W., Wolfe, J., & Cotter, R. V. (1999). A test bank for measuring total enterprise simulation learning. *Developments in Business Simulation & Experiential Exercises*, *25*, 82-92.

Gosen, J., Washbush, J., & Scott, T. W. (2000). Initial data on a test bank assessing total enterprise simulation learning. *Developments in Business Simulation & Experiential Exercises*, *26*, 166-171.

Gosenpud, J. (1986). Personality variables on group cohesion, team participation and total learning. *Developments in Business Simulation & Experiential Exercises*, *13*, 115-118.

Gosenpud, J. (1990). Evaluation of experiential learning. In J. Gentry (Ed.), *Guide to business simulation and experiential learning* (pp. 279-300). New Brunswick, NJ/London: Nichols/GP.

Graf, L. A., & Kellogg, C. E. (1986). Justifications for and problems in developing and using computerized experiential activities. *Developments in Business Simulation & Experiential Exercises*, *13*, 119-121.

Gray, C. F. (1972). *Performance as a criterion variable in measuring business gaming success: An experiment with a multiple objective performance model*. Paper presented at the Southwestern American Institute of Decision Sciences.

Henshaw, R. C., & Jackson, J. R. (1984). *THE EXECUTIVE GAME*. Burr Ridge, IL: Irwin.

Herbert, E. P., & Landin, D. (1996). Practice schedule effects on the performance and learning of low- and high-skilled students: An applied study. *Research Quarterly for Exercise and Sport*, *67*, 52-58.

Hergert, M., & Hergert, R. (1990). Factors affecting student perceptions of learning in a business policy game. *Developments in Business Simulation & Experiential Exercises*, *17*, 92-96.

Herz, B., & Merz, W. (1998). Experiential learning and the effectiveness of economic simulation games. *Simulation & Gaming*, *29*, 238-250.

Hirsch, E. D., Jr. (2002). Classroom research and cargo cults. *Policy Review*. Retrieved January 6, 2003, from http://www.policyreview.org/OCT02?hirsh_print.html

James, P. (2000). The influence of a period of environment oriented work on students' perception of their learning style. *Environmental Education Research*, *6*, 157-165.

Keys, B., & Wolfe, J. (1990). The role of management games and simulations in education and research. *Journal of Management*, *16*, 307-336.

Keys, J. B. (1986, March). *A proposal for cooperation between ABSEL and Georgia Southern's Center for Advancement of Business Simulation and Experiential Learning*. Paper presented at the Association for Business Simulation and Experiential Learning, Reno, NV.

Kline, D. S. (1986). Management competency models and the Life-Long Learning Project: What role for ABSEL. *Developments in Business Simulation & Experiential Exercises*, *13*, 122-125.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals, Handbook II: The affective domain*. New York: David McKay.

Leonard, T. L., & Leonard, N. J. (1995). Graduates' views on the use of computer simulation games versus cases as pedagogical tools. *Developments in Business Simulation & Experiential Exercises*, *22*, 83-87.

Lewin, K. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.
Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.
Manoque, M., Brown, G. A., Nattress, B. R., & Fox, K. (1999). Improving student learning in root canal treatment using self-assessment. *International Endodontic Journal*, *32*, 397-405.
McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale, NJ: Lawrence Erlbaum.
McHaney, R., White, D., & Heilman, G. E. (2002). Simulation project success and failure: Survey findings. *Simulation & Gaming*, *33*, 49-66.
McKenney, J. L., & Dill, W. R. (1966). Influences on learning in simulation games. *American Behavioral Scientist*, *10*, 28-32.
Nichols, J. D., & Steffy, B. E. (1999). The evaluation of success in an alternative learning programme: Motivational impact vs. completion rate. *Educational Review*, *51*, 207-219.
Norris, D. R. (1986). External validity of business games. *Developments in Business Simulation & Experiential Exercises*, *13*, 126-129.
Peach, E. B., & Platt, R. G. (2000). Total enterprise simulation and optimizing the decision set: Assessing student learning across decision periods. *Developments in Business Simulation & Experiential Exercises*, *27*, 242-245.
Premi, J., & Shannon, S. I. (1993). Randomized controlled trial of a combined video-workbook. Educational Program for CME. *Education in the Information Age*, *10*(October Suppl.), S13-S15.
Premi, J., & Shannon, S. I. (2001). Randomized controlled trial of an educational program for individualized learning. *Journal of Continuing Education in the Health Professions*, *17*, 245-249.
Raia, A. P. (1966). A study of the educational value of management games. *Journal of Business*, *39*, 339-352.
Rocha, C. J. (2000). Evaluating experiential teaching methods in a policy practice course: The case for service learning to increase political participation. *Journal of Work Education*, *36*, 53-63.
Schreier, J. (1981). Research questions for cases. *Developments in Business Simulation & Experiential Exercises*, *8*, 71.
Scott, T. W., Strickland, A. J., Hofmeister, D. L., & Thompson, M. D. (1992). *MICROMATIC: A management simulation*. Boston: Houghton Mifflin.
Smith, C. A., Strand, S. E., & Bunting, C. J. (2002). The influence of challenge course participation on moral and ethical reasoning. *Journal of Experiential Education*, *25*, 278-280.
Specht, L. B., & Sandlin, P. K. (1991). A comparison of the effects of experiential learning activities and traditional lecture classes. *Developments in Business Simulation & Experiential Exercises*, *17*, 214.
Thorngate, W., & Carroll, B. (1987). Why the best person rarely wins. *Simulation & Games*, *18*, 299-320.
Tompson, G. H., & Dass, P. (2000). Improving students' self-efficacy. *Simulation & Gaming*, *31*, 22-41.
Vance, S. C., & Gray, C. F. (1967). Use of a performance evaluation model for research in business gaming. *Academy of Management Journal*, *10*, 27-37.
Washbush, J., & Gosen, J. (2001). Learning in total enterprise simulations. *Simulation & Gaming*, *32*, 281-296.
Washbush, J. B., & Gosenpud, J. (1991). Student attitudes about policy course simulations. *Developments in Business Simulation & Experiential Exercises*, *18*, 105-110.
Wellington, W. J., & Faria, A. J. (1991). An investigation of the relationship between simulation play, performance level and recency of play on exam scores. *Developments in Business Simulation & Experiential Exercises*, *18*, 111-115.
Wellington, W. J., & Faria, A. J. (1992). An investigation of the relationship between team cohesion, player attitude, performance attitude on player performance. *Developments in Business Simulation & Experiential Exercises*, *19*, 184-189.
Wellington, W. J., & Faria, A. J. (2001). An investigation of the awareness attained in a simple business simulation game. *Developments in Business Simulation & Experiential Exercises*, *28*, 239-249.
Wheatley, W. J., Hornaday, R. W., & Hunt, T. G. (1988). Developing strategic management goal setting skills. *Simulation & Games*, *19*, 173-185.
White, C. S., & Von Riesen, R. D. (1992). Computer management and some correlates of students' satisfaction. *Developments in Business Simulation & Experiential Exercises*, *19*, 225.
Whiteley, T. R., & Faria, A. J. (1989). A study of the relationship between student final exam performance and simulation game participation. *Simulation & Games*, *20*, 44-64.

Wolfe, J. (1976). Correlates and measures of the external validity of computer-based business policy decision making environments. *Developments in Business Simulation & Experiential Exercises*, *7*, 411-438.
Wolfe, J. (1986). The ABSEL research consortium: Preliminary steps and position papers. *Developments in Business Simulation & Experiential Exercises*, *13*, 96.
Wolfe, J. (1990). The evaluation of computer-based simulation games. In J. Gentry (Ed.), *Guide to business simulation and experiential learning* (pp. 279-300). New Brunswick, NJ/London: Nichols/GP.
Wolfe, J. (1997). The effectiveness of business games in strategic management course work. *Simulation & Gaming*, *38*, 360-376.
Wolfe, J., & Crookall, D. (1998). Developing a scientific knowledge of simulation/gaming. *Simulation & Gaming*, *29*, 7-19.
Wolfe, J., & Guth, G. (1975). The case approach vs. gaming in the teaching of business policy: An experimental evaluation. *Journal of Business*, *48*, 349-364.
Wolfe, J., & Luethge, D. J. (2003). The impact of involvement on performance in business simulations: An examination of Goosen's "know little" decision making thesis. *Journal of Education for Business*, 69-74.
Wolfe, J., & Roberts, C. R. (1986). The external validity of a business management game: A five-year longitudinal study. *Simulation & Games*, *17*, 45-59.
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.
Wolfe, J., & Rogé, J. (1997). Computerized general management games: The effectiveness of business games in strategic management course work. *Simulation & Gaming*, *28*, 423-441.
Zalatan, K. A., & Mayer, D. F. (1999). Developing a learning culture: Assessing changes in student performance and perception. *Developments in Business Simulation & Experiential Exercises*, *26*, 45-51.

*Jerry Gosen received his PhD in organizational behavior in 1974 from Case Western Reserve University. He is currently a professor of management at the University of Wisconsin–Whitewater and is a past president and fellow of the Association for Business Simulation and Experiential Learning (ABSEL).*

*John Washbush received his PhD in educational administration in 1975 from Marquette University. He is currently a professor of management at the University of Wisconsin–Whitewater.*

*ADDRESSES: JG: Management Department, University of Wisconsin–Whitewater, Whitewater, WI 53190, USA; telephone: +1 (262) 472-3956 (w), +1 (847) 247-8247 (h); fax: +1 (262) 472-4263; e-mail: gosenpuj@uww.edu. JW: Management Department, University of Wisconsin–Whitewater, Whitewater, WI 53190, USA; telephone: +1 (262) 472-3902 (w), +1 (608) 271-3689 (h); fax: +1 (262) 472-4263; e-mail: washbusj@uww.edu.*