COMMENTS ON THE PERCEPTION, IDENTIFICATION, AND MEASUREMENT OF
LEARNING FROM SIMULATION GAMES

Remarks for

Third National Conference on
Business Simulation and Experiential Learning

Knoxville, Tennessee
April, 1976

Prepared by:

Joseph Wolfe
Associate Professor of Management
University of Illinois, Chicago Circle

Submitted to:

Bernard Keys
Assistant Dean
Tennessee Technological University
Cookeville, Tennessee
The Design of Studies on the Comparative Effects of Gaming versus Traditional Teaching Methods

Although a number of evaluative typologies exist, the systems of James [4] or Suchman [12] are particularly useful when considering the gaming literature. In their systems, the success or failure of a gaming application is judged by the (1) effort expended by the applicator, (2) obtained performance outcomes, (3) adequacy of the application’s performance, (4) relative efficiency of the application when compared to alternative teaching methods, and (5) process and affects through which results are presumed to occur. While the early gaming literature abounded in reportorial and descriptive studies of the first category and there has been a continuing interest in studies of the last category (termed an epistemological approach by Keys [51), this short note has been delimited to those studies that are basically comparative in nature and are primarily interested in the efficacy of one pedagogical approach over some other pedagogical approach.

For studies that are basically comparative in nature, Campbell and Stanley’s [1] posttest-only control group design should serve as the barest minimum. Unfortunately, because random assignment to the experimental and control groups is rarely achieved in practice, more rigorous designs must be employed such as the classical pretest-posttest control group design. Even this design, however, has faults as it fails to control for the interaction of testing and the treatment, and it is weak in determining the effects of the (1) interaction between being part of the study and the treatment (the “elite” aspects of the Hawthorne effect), and
(2) reactive arrangements. This latter problem is an especially invalidating factor in the McKenney and Dill study on the effects of team groupings based on grade levels and course performance and performance in a simulated industry. As the authors noted, “because high performers and over-achievers tended to be people recognized within the class as discussion leaders in other courses, the class knew that the team assignments were not random. There was some resentment toward the game administrators for what was perceived as an obvious bias in the groupings [8:31].” Another example would be Rowland and Gardner [10] where students were required to make their decisions within a three-tiered hierarchy even though it had to be apparent to most players that self-selected hierarchies or an organic structure would have been more comfortable to them and/or more effective.

The most rigorous design application used thus far has been the separate-sample pretest-posttest control group design as modified by Wolfe [15] [17] This design, in its pure form, controls for all sources of invalidity except for the interaction of selection and maturation. But Wolfe did not obtain the pure form. Row (1) was not created due to a lack of subjects and an inability to create the placebo (X) in column (c). It is also doubtful that randomization, R, was obtained in rows (4) and (5), and it is certain that R did not exist in (2) as students elected to join the experimental program.
The space devoted thus far in this note should make it apparent that the research’s design is important, that rigorous research designs are available, but that the perfect theoretical implementation has not been obtained nor is it likely to ever occur. What is important then is that gaming evaluations create the best design possible and apply it conscientiously. True randomization will never be achieved when students are given any choice over either course sections or course drops. The test instruments used will never completely capture all the learning, both cognitive and affective, that occurs in the experiential learning environment. And the learning sources from the simulation, the group, and the individual’s past and present, will never be definitively located.

The test instruments, either case analyses or essay-type examinations, Used by McKenney [6] [7], Raia [9], or Wolfe [14] [17] [18] all suffer from subjectivity and a lack of standardization. Thus it is unlikely their results could be obtained again even if their studies could be replicated. When objective, multiple-choice examinations have been employed, it is apparent that they can incompletely or incorrectly capture what students learn from a simulation game. In Thompson and Schrieber [13:167], the subjects themselves “...expressed doubt that the tests [used] could measure what they were intended to measure.” As for the sources of learning in a game experience, the research designs most often employ a student post-hoc reaction questionnaire [3] [15] [17]; less often, faculty observations of the learning process are employed [2] [3] [7]. Very rarely are the supposed gaming benefits tested against quantifiably behavioral or objectively observed and
recorded measures. Schriesheim and Yaney [11] are an example of what can be done in this important research area.

Accordingly, as important as it is to create the proper evaluation research design, it must be recognized that many other invalidating and confounding factors enter the research situation. These factors are embraced in the physical and academic setting, the test instruments used, the codification of responses, and instructor and student biases. Too often in our search for the correct research design do we fail to recognize these other contingencies which can either foil the perfect design or lead to false conclusions regarding the value of games and experiential learning environments.
REFERENCES


