Developments In Business Simulation & Experiential Exercises, Volume 22, 1995

THE PROBLEM OF DETERMINING AN "INDIVIDUALIZED" SIMULATION'S VALIDITY AS AN ASSESSMENT TOOL

Philip H. Anderson, University of St Thomas
Leigh Lawton, University of St Thomas

ABSTRACT

An "individualized" version of a competitive, total enterprise (TE) simulation exercise was tested as a potential instrument for measuring the skills developed on a group-managed simulation. No significant relationship was found between a subject's performance on the "individualized" version of the simulation and his or her performance on the group-managed version of the simulation. Further, no relationship was found between performance on the "individualized" version of the simulation and a set of independent variables including: a peer group assessment of the subject's strategic management skills; a self-assessment of managerial skills; the grade received on a case study write-up; the grade received for class participation during the course; and overall GPA level. Only the subject's business GPA showed a significant relationship with performance on the "individualized" version of the simulation. The need to develop learning goals and, especially, valid and reliable methods for measuring learning is discussed as the key for evaluating a simulation's assessment capabilities.

INTRODUCTION

One of the continuing problems associated with the use of groups in business simulations is the difficulty of assessing the performance of an individual within the group (Wolfe and Box, 1988). The predominant pedagogy used for business simulation exercises is a group format; i.e., teams competing against other teams (Anderson and Lawton, 1992b). The use of groups is effective for *creating* a learning environment that confronts students with the group dynamics involved in managing a competitive enterprise. However, their use limits the instructor's ability to *assess* individual contribution to, and comprehension of, the exercise (Burns, Gentry, and Wolfe, 1990).

Assessing the learning that occurs on a business simulation exercise has been a long-standing issue in the literature (Greenlaw and Wyman, 1973; Keys, 1976; Wolfe, 1981; Wolfe, 1985; Whiteley and Faria, 1989; Burns, Gentry, and Wolfe, 1990; Wolfe, 1990; Gosenpud, 1990; Gosenpud and Washbush, 1993; Gosenpud and Washbush, 1994). A resolution to the conundrum of "What is *learned* on a simulation exercise and how do we measure it?" has not yet been achieved (Gosenpud and Washbush, 1994). The lack of an appropriate means for assessing learning interferes with the instructor's ability to accurately and fairly assess what a student has learned from the simulation exercise. This has consequent effects on accurate and fair grade assignment.

Over the years, multiple methodologies have been utilized in an attempt to assess individual performance in a simulation exercise. These include paper and pencil tests on simulation rules and procedures, peer evaluations of individual contributions, and written essays analyzing company performance (Anderson and Lawton, 1992b; Gosenpud and Washbush, 1994). Attention also has been given to measuring students' *perceptions,* as opposed to objective measures, of what they learned (e.g.; Schellenberger, et *a!,* 1989). Unfortunately, none of these methods provides a direct test of an individual's ability to learn how to manage a simulation exercise on *his or her own.*

Further, most of these methods have focused on a type of learning that is at a different level from that at which simulations operate. Bloom's Taxonomy of Learning (Bloom et al, 1956) classifies learning into six levels. These levels are arranged in a hierarchical order to reflect progressively higher levels of learning. They are, in ascending order, basic knowledge, comprehension, application, analysis, objective synthesis, and objective evaluation. Earlier research has shown that paper and pencil tests tend to fall at the lower end of Bloom's Taxonomy, while performance on the simulation lies at the upper end (Anderson and Lawton, 1988). Consequently, paper and pencil tests, alone, cannot adequately assess the type of learning, which occurs from participating in a simulation exercise (Wellington and Faria, 1991). At present, there are no objective measures, which assess learning at the higher levels of Bloom's Taxonomy.

The absence of objective measures of learning has led to a reliance on financial results to evaluate student performance at all levels of learning (Anderson and Lawton, 1992b). However, relying on the group's financial performance on the simulation to determine

individual grades can unfairly reward or penalize individual members of the group; as they are "carried" or "held back" by other members of the group, respectively (Gosenpud and Washbush, 1994). Further, the relationship between financial performance and other measures of student learning has been found to be weak or non-existent (Anderson and Lawton, 1992a). Therefore, reliance on team financial performance as a proxy for individual learning appears to be fraught with error.

The purpose of this study was to assess the relationship between performance on an "individualized" version of a group-managed, competitive, TE simulation exercise and other measures of performance. The goal was to determine whether the individually-operated version of a group-based simulation can serve as a tool for assessing the individual learning acquired from participating in that group-based simulation.

RESEARCH METHODOLOGY

The Subjects and the Course

Subjects for the study were seniors in their final year of study at a medium-sized, midwestern university. All were members of one of two sections of a business policy course conducted using a total enterprise simulation (described below) as a major component of the course pedagogy. The profile of the student was that of a typical, traditional college senior. All were majoring in various fields of business administration.

The Simulation

The simulation used was *Micromatic: A Strategic Management Simulation,* 2nd ed by Scott, et a! (1992). *Micromatic* is a moderately complex simulation. Each decision set requires approximately 60 decisions in the areas of marketing, production, and finance. Each decision represents a three-month period.

The Group-based Model. Students were first exposed to the simulation as a member of a group. Each group was in competition with the other groups in its section of the course. Twelve simulated quarters of operation were run. The group-based portion of the simulation exercise comprised about one-half of the course in terms of pedagogy and time, and determined 25% of the student's final grade.

The Individualized Model. After completion of the group-based simulation exercise, each student was given an "individualized" version of the group-based simulation. From the student's perspective, the "individualized" version is identical to the group-based version. The student decisions and decision screens are exactly the same for both models. The differences between the two models occur when the student company's decisions are processed. In the group-based version, each student team turns its decision set into the instructor for processing. The instructor then submits the decisions for all teams in an industry to the computer. The results achieved by each team are dependent upon the actions of the other student teams. In the individual version, a modified processing program is included on the student's disk. This allows the student to process his/her quarter's decisions without having to submit them to the instructor. Further, the competitors are not other student teams, but are "semi-intelligent" competitors managed by the computer. These computer competitors react to the decisions made by the student run company. While they act rationally, the computer competitors are not all-seeing and all-knowing, "perfect competitors". This allows the student with a well-defined and well-executed strategy to beat the competition.

The Simulation Performance Index (SPI)

Micromatic uses seven factors to determine current quarter and game-to-date rankings of company performance. These factors are sales revenues, net income, earnings per share, return on sales, return on assets, return on equity, and stock price. The percentage weights assigned to these factors was the same for both the group and the individual exercises.

The Dependent Variable

The dependent variable measured was the simulation performance index for the individual exercise (i.e.; the Individual SPI).

The Independent Variables

Seven independent variables were measured to check for relationships with the individual simulation performance measure. These were: the simulation performance index for the group exercise (i.e.; the Group SPI); an assessment of an individual's strategic management skills by other members of the simulation group; a self assessment of personal managerial skills; the grade received on a case study write-up; the grade received for class participation during the course; the subject's business GPA; and the subject's overall GPA.

A modified version of Stumpfs Strategic Management Skills Questionnaire (i.e.; the SMSQ) was used to measure an individual's strategic management skills. Each member of the group simulation exercise evaluated his/her teammates. Stumpf identified six strategic management skills; knowing the business, managing subunit rivalry, finding and overcoming problems, staying on strategy, being an entrepreneurial force, and accommodating adversity (Stumpf, 1988). Stumpf reports internal consistency estimates of reliability medians of .74 and validity coefficient medians of .57 for his 20-item instrument.

Each student rated his or her personal managerial skills using the Managerial Effectiveness Profile System (i.e.; the MEPS) developed by Human Synergistics (1983). A modified version of the MEPS (i.e.; the MMEPS) was designed for use with students (Anderson and Lawton, 1990). The MMEPS has 71 items that form fourteen scales such as planning effectively, managing conflict, making decisions, and demonstrating commitment. The Cronbach alpha for the MMEPS score, using a pool of 125 undergraduate students, was .95.

Research Hypotheses

The hypotheses tested in this study were:

- H1: The Individual SPI will be positively related to the Group SPI.
- H2: The Individual SPI will be positively related to the SMSQ score.
- H3: The Individual SPI will be positively related to the MMEPS score.
- H4: The Individual SPI will be positively related to the case write-up grade.
- H5: The Individual SPI will be positively related to the class participation grade.
- H6: The Individual SPI will be positively related to the business GPA.
- H7: The Individual SPI will be positively related to the overall GPA.

RESULTS

A Pearson Correlation analysis was conducted for all of the variables to test for significant associations. The analysis found no significant relationship between performance on the "individualized" simulation exercise and performance on the group-based simulation (See Table 1). The only variable significantly related with the Individual SPI was the Business GPA. None of the other variables measured (i.e.; SMSQ, MMEPS, the case write-up or class participation, overall GPA) were significantly related with the individual SRI. Only Hypotheses H6 received any support.

Further, there were few significant associations between the independent variables measured. All but one of these significant associations were between either the business or overall GPA and another variable. These relationships were not unexpected. The lack of any significant relationship between the GPA measures and the Group SPI could be the result of the group effects discussed earlier. (This provides support for the need for techniques that allow assessment of individual learning.) With one exception, none of the paper and pencil tests or the instructor assigned grades showed any relationship with either the individual or group simulation performance indexes or with each other.

To further examine the interrelationships among the variables, all scores for students in the middle half of the "individualized" simulation were dropped from the analysis and the correlations were rerun. The goal of this analysis was to see whether any useful relationships would be evident using only the best 25% and the worst 25% of the performers on the "individualized" simulation. We reasoned that, while the measures used in this study may not exhibit fine degrees of discrimination, if any relationships exist, they should be evident when we look at the very best performers against the very worst performers. The results of this analysis are shown in Table 2.

Again, as in Table 1, no significant relationships were found for *any* of the variables, except for the Business and Overall GPAs. No relationship was found between performance on the group simulation exercise and the best and worst performers on the individual simulation exercise. However, this relationship was approaching significance (p =06). The small sample size (n=19) which resulted from eliminating the middle 50% performers may have hindered significance from being reached. Regardless of the near significance between the Individual and Group SPIss, the absence of significant relationships between the variables *after* limiting analysis to only the best and the worst performers on the individual simulation was disturbing.

Table 1
Correlation Analysis - All Subjects

	Individual SPI	Group SPI	SMSQ	MMEPS	Case Write-up	Class Participation	Business GPA
Group SPI	.25 NS						
SMSQ	.11 NS	.01 NS					
MMEPS	-.21 NS	.11 NS	.14 NS				
Case Write-up	-.04 NS	.10 NS	.18 NS	.46 .02			
Class Participation	.16 NS	.23 NS	.30 NS	.01 NS	.35 NS		
Business GPA	.42 .01	.14 NS	.44 .01	.03 NS	.51 .01	.59 .01	
Overall GPA	.29 NS	.19 NS	.43 .01	.23 NS	.61 .01	.58 .01	.93 .00

top number = r
Bottom number = p-value (NS = if not significant at p = .05)

Table 2
Correlation Analysis - Top & Bottom 25%

	Individual SPI	Group SPI	SMSQ	MMEPS	Case Write-up	Class Participation	Business GPA
Group SPI	.41 NS						
SMSQ	.30 NS	.06 NS					
MMEPS	-.25 NS	.33 NS	-.01 NS				
Case Write-up	.06 NS	.06 NS	-.06 NS	.49 NS			
Class Participation	.30 NS	.26 NS	.31 NS	-.08 NS	.18 NS		
Business GPA	.60 .01	.09 NS	.56 .01	.01 NS	.34 NS	.53 .01	
Overall GPA	.44 .03	.20 NS	.47 .03	.20 NS	.54 .01	.50 .02	.91 .00

top number = r
Bottom number = p-value (NS = if not significant at p = .05)

DISCUSSION AND CONCLUSIONS

A number of issues need to be recognized before drawing any conclusions from the results of this study. First, the potential pitfalls of self-assessment instruments such as the MMEPS used in this study, are quite clear. Students may be unable to accurately assess their own strengths and weaknesses. Furthermore, even if they do perceive weaknesses, they may be unwilling to admit them to others --especially to the course instructor. While the SMSQ is based, not on self-assessment, but on the perceptions of a student's teammates, it too is subject to some obvious problems. One group member's evaluation of another may be colored by the affective relationship between the two. In addition, if a group member is shy or undermotivated, his or her evaluation may be a better reflection of his or her contribution to the group rather than of managerial ability. Since both MMEPS and SMSQ are designed to measure the same

construct -- managerial skills -- we would expect to find a high degree of association between the two measures. The fact that there is virtually no correlation between MMEPS and SMSQ scores is an indication that one or both of the measures lack validity.

Second, the group simulation SRI is unlikely to be a particularly good measure of individual learning or ability. The ineffectiveness of weak students may be disguised by contributions of more capable teammates; the efforts of strong performers may be thwarted by less capable teammates; and the combined competency of the members of a group may differ from the capabilities of the individual team members operating alone. The absence of a significant relationship between the Group SPI and either the Business or Overall GPAs lends credence to this view and further magnifies the need for a way to measure individual learning. It is generally assumed that GPA reflects a student's ability to demonstrate learning (an assumption beyond the scope of this paper). *!f* this assumption is correct, this study's results suggest that group dynamics inhibited many of the best students from being the best performers, while "sheltering" the worst students from exhibiting poor performance. Recognition of the inability to accurately assess individual performance in a group setting was the reason for attempting to use the "individualized" simulation as a measurement tool. However, without valid methods to measure learning, it is impossible to determine the learning acquired in, or the assessment capabilities of, any simulation model.

Third, class participation is almost certainly not a completely valid measure of higher-level thinking skills. Evaluations will be contaminated by how articulate the student is and how out-spoken he or she is. The case write-up may be the measure most similar to the simulation. It is an individual exercise and is not likely to be as susceptible to the subjective judgment of the instructor as is class participation. Even so, the results of this study found no relationship between the case write-up and performance on *either* the individualized or the group simulation exercises.

Fourth, the lack of significant relationships among the independent variables is a strong sign that the variables either lack validity or are measuring different things. Thus, failure to find significant relationships between performance on the individualized simulation and the various independent variables may be less of an indictment of the individualized simulation as an assessment tool than it is a reflection that none of the other measurements are meaningful. It is conceivable that the scores on the "individualized" simulation are the *only* valid measures we have. Unfortunately, having no known, valid measure of the higher levels of learning, it is impossible to establish the construct validity of the "individualized" simulation scores.

We had hoped to find that the scores on the "individualized" simulation would be related to scores of the independent variables. If we had, we could have advocated the use of the individualized simulation as an assessment tool and, on an exploratory basis, as a proxy for measuring learning. Since this did not occur, what meaning can we draw from our results? The disconcerting truth is that there is an abysmal absence of valid and reliable tools and techniques, which can measure learning at the higher levels of Bloom's Taxonomy. Simple paper and pencil tests cannot adequately measure an individual's ability to synthesize, which is a key determinant of success on the simulation exercise. Without better measures of learning at the upper levels of Bloom's hierarchy, we simply can't validate the individualized simulation - nor can we validate the group simulation exercise.

After decades of research on simulation exercises, we still cannot provide objective (versus anecdotal) support for answers to questions like Does participating in simulation exercise produce learning? If so, what kinds of learning and how do we measure it? If not, of what value are simulations? Could the time spent on simulation exercises be used more effectively or productively if directed toward other pedagogies? Neither is there a consensus regarding the questions of: What knowledge, skills, and attitudes do we want students to learn from a simulation exercise? Which simulation measures (e.g., financial performance), if any, reflect these learning goals?

The problem of how to measure Bloom's upper levels of learning is of paramount importance to simulation researchers. This is where the focus of simulation research must now be directed. We must also delineate what "learnings" we want students to acquire as the result of participation in a simulation exercise. Then, we must link the measures of learning with the goals of learning. This line of research must be pursued if we are to make significant progress in understanding the value of simulation exercises. Simulations appear to hold considerable promise for being able to measure changes in *abilities* versus *attitudes.* The lack of direct linkages between attitudes and abilities has been demonstrated elsewhere (see, for example, Schumann, Scott, and Anderson, i994). It is the teaching of abilities at the higher levels of Bloom's Taxonomy that has been the charge for

instructors using simulations. We need to objectively assess the effectiveness of this pedagogy if we are to maintain our credibility as educators.

Many of us (including these authors) shy away from using the phrase "simulation *game",* preferring "simulation *exercise".* Perhaps Freud is at work again. We may believe "Emperor Simulation" is dressed in educational sartorial splendor. We need to objectively determine whether that is, in fact, true.

REFERENCES

Anderson, P.H. and L. Lawton (1988), "Assessing Student Performance on a Business Simulation Exercise." *Developments in Business Simulations and Experiential Exercises,* 15, 241-244

Anderson, P.H. and L. Lawton (1990), "Measuring the Learning Outcomes of Management Training Activities", *Association of Management Conference.*

Anderson, P.H. and L. Lawton (1992a), "The Relationship Between Financial Performance and other Measures of Learning on a Simulation Exercise", *Simulation & Gaming,* 23 (3), 326-340.

Anderson, P.H. and L. Lawton (1992b), "A Survey of Methods Used for Evaluating Student Performance on Business Simulations", *Simulation & Gaming,* 23, 490-498.

Bloom, B.S., M.D. Englehart, E.D. Furst, W.H. Hill, and D.R. Krathwohl (1959), *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain,* New York: David McKay Company, Inc.

Burns, A.C., J.W. Gentry, and J. Wolfe (i990), "A Cornucopia of Considerations in Evaluating the Effectiveness of Experiential Pedagogies," in *Guide to Business Gaming and Experiential Learning,* J.W. Gentry (ed.), New York: Nichols/GP Publishing, 253-278.

Gosenpud, J. (1990), "Evaluation of Experiential Learning," in *Guide to Business Gaming and Experiential Learning,* J.W. Gentry (ed.), New York: Nichols/GP Publishing, 301-329.

Gosenpud, J. and J. Washbush (1993), "The Relationship Between Total Enterprise Simulation Performance and Learning." *Developments in Business Simulations and Experiential Exercises,* 20, 141.

Gosenpud, J. and J. Washbush (1994), "Simulation Performance and Learning Revisited" *Developments in Business Simulations and Experiential Exercises,* 21, 83-86.

Greenlaw, P.S. and F.P. Wyman (1973), "The Teaching Effectiveness of Games in Collegiate Business Courses." *Simulation and Games,* 4, 259-293.

Keys, B. (1976), "A Review of Learning Research in Business Gaming." In B.H. Sord (ed.), *Computer Simulations and Learning Theory,* 173-184.

Schellenberger, R.E., J.A. Hill, and RB Keusch (i989), "An Exploratory Study of the Effect of Strategic Emphasis in Management Games on Attitudes, Interest, and Learning in the Business Policy Course." *Developments in Business Simulations and Experiential Exercises,* 16, 178.

Schumann, P.J., T.W. Scott, and PH. Anderson (1994), "Using a Business Simulation to Study the Determinants of Ethical Behavior", *Developments in Business Simulation and Experiential Exercises,* 22, 90-95.

Scott, T.W., A.J. Strickland, D.L. Hofmeister, and M.D. Thompson (1992), *Micromatic: A Strategic Management Simulation,* 2nd ed. Boston, Houghton Mifflin.

Stumpf, S.A. (1988), "Using the Nest Generation of Assessment Center Technology for Skill Diagnosis", in J.W. Jones, B.D Steffy, and D.W. Bray (Eds), *Applying Psychology in Business: The Manager's Handbook.* Lexington, MA: Lexington Books.

Teach, R.D. (i987), "Profits: The False Profit." *Developments in Business Simulations and Experiential Exercises,* 14, 205-207 (1983), *Managerial Effectiveness Profile System,* Human Synergistics. Plymouth, Mi.

Wellington, W.J. and A.J. Faria (1991), "An Investigation of the Relationship Between Simulation Play, Performance Level, and Recency of Play on Exam Scores", *Developments in Business Simulations and Experiential Exercises,* 18, 177.

Whiteley, T.R. and A.J. Faria (1989), "A Study of the Relationship between Student Final Exam Performance and Simulation Game Participation." *Developments in Business Simulations and Experiential Exercises,* 16, 78-82

Wolfe, J. (1981), "Research on the Learning Effectiveness of Business Simulation Games – A Review of the State of the Art." *Developments in Business Simulations and Experiential Exercises,* 9, 72.

Wolfe, J. (1985), "The Teaching Effectiveness of Games in Games in Collegiate Business Courses:A 1973-83 Update." *Simulation and Games,* 16, 251 -288.

Wolfe, J. (1990), "The Evaluation of Computer-based Business Games: Methodology, Findings, and Future Needs," in *Guide to Business Gaming and Experiential Learning,* J.W. Gentry (ed.), New York: Nichols/GP Publishing, 279-300.

Wolfe, J. and TM Box (i988), "Team Cohesion Effects on Business Game Performance," *Simulation and Games,* 19, 82-89.