The fields of operational gaming and micro-economic modeling deal with the symbolic representation of real world systems or events. In business gaming the simuland is the individual firm or one of its functional areas while the program that simulates the system would be the simulator (Stanislaw, 1986). Although the ideal for these two fields is isomorphism a number of factors frustrate the acquisition of this ideal; the lack of time, and developmental monies, the presence of imperfect theory, bounded programming skills and the natural confines of the programming languages employed, and hardware constraints.

Given these realities validity has more recently come to be judged in a comparative or task related sense; the validity of a simulator is a function of what designer wants the model to accomplish (Schrank and Holt, 1967; Stanislaw, 1986; Van Horn, 1971). Using the concept of event validity Mehrez, Reichel and Olami (1987) have evaluated a derivative of the Carnegie Tech Management Game (Cohen, Dill, Kuehn and Winters, 1964) where the game recreated Israel’s detergent industry. The performance of the simulation’s Deterclean Company was compared to the performance of the real world Mecca Chemicals Ltd. Based on output measures the performance results were quite similar, especially for the first year of operations. Accordingly, the authors felt the simulation possessed high face validity although no tests were made of the teaching value of this apparent validity.

In a more theory testing sense, numerous studies by Gold/Pray (Gold and Pray, 1983; Gold and Pray, 1984; Pray and Gold, 1982; Pray and Gold, 1984) have examined the hypothesis validity (Hermann, 1967) possessed by a number of business games. As examples of these efforts Pray and Gold (1982) found that the demand functions used in four business games featured illogical and unstable demand functions beyond certain ranges, and employed algorithms which did not employ economic theory’s most current thinking (Gold and Pray, 1983). From a learning perspective regarding hypothesis validity, Wolfe and Teach (1987) found that the demand function employed in The Executive Game (Henshaw and Jackson, 1984) “worked” and made playing to the simulation’s algorithms or otherwise “cracking” the game almost impossible, but that the demand function also employed a number of constants possessing little or no economic sense within the frame description outlined in the player’s manual.

Educational Assessment

The empirically based evaluation literature is quite barren although it abounds in testable theories and conceptual speculations. As observed by Wolfe (1985) no objective research has been conducted on the relationship between a game’s degree of realism and the amount of learning that accompanies such realism. Regarding subjectively assessed learning levels; however,
Developments in Business Simulation & Experiential Exercises, Volume 16, 1989

three different games of increasing levels of face complexity have been evaluated. As part of a larger research effort (Dittrich, 1975) that created a game realism instrument, Dittrich (1976) compared the perceived realism of three general management games.

The Executive Game (Henshaw and Jackson, 1972), INTOP (Thorelli, Graves and Hidwells, 1964) and The Business Management Laboratory (Jensen and Cherrington, 1973) were considered non-significantly different regarding their overall degree of realism even though their complexity levels were vastly different when objectively measured by such external game features as the number of products offered, marketing territories and financial sources available to each firm, as well as the number of decisions allowed per round. Within the functional areas of marketing and production The Business Management Laboratory (BML) was considered the most realistic while INTOP possessed the most realistic financial function. Participants felt the BML made the greatest contribution to their subjectively assessed learning level even though the game occupied a middle position regarding game complexity.

Although a game’s complexity and its perceived degree of realism may not be synonymous entities, the interaction between game complexity and objectively-assessed learning has been primarily investigated by Thompson and Hansen (1965), Rhenman (1961), and Thompson and Schrieber (1966), and by Raia (1966), and Wolfe (1978) with a partial replication by Butler, Pray and Strang (1979). The work by Thompson and associates did not comparatively test games of increasing complexities for their teaching efficacies but instead objectively measured pre/post-learning levels obtained by different games, some of which were “simple,” while others were “complex.” Unfortunately the simulations were not identified nor were the instruments administered in a uniform manner across the various schools employed in the study.

In a tightly-controlled study Raia concluded that increased game complexity did not lead to greater business policy knowledge or increased motivation within Models I and II of MANSYM (Schellenberger, 1965). Wolfe found from a very wide range of three increasingly complex games, however, that the most complex simulation was superior in all three learning categories examined. The most complex game also produced the greatest degree of decision-making comprehensiveness as well as the greatest degree of self-assessed intellectual challenge and excitement while simultaneously producing the greatest degree of student mortality. A partial replication Butler, Pray and Strang (1979) did not objectively measure learning levels, but it was established that an intermediately complex game (Pray and Strang, 1977) was equally challenging and exciting for both senior and freshman/sophomore level students although higher attrition rates were found for the latter group. It was concluded that factors other than complexity might be more influential regarding a game’s teaching qualities and challenge levels.

Many business educators have been more concerned with administrative and face validity matters than with a particular game’s algorithmic validity as they outline the simulation’s need for plausibility, credibility, and verisimilitude all in the name of a player’s interest and motivation to participate in the environment created by the simulation. Graham and Gray (1969) employed the concept of game fidelity while simultaneously dealing with a game’s attitudinal effects of teaching and learning. For them a game should possess high fidelity and interactive complexity so that it presents an environment rich in learning potential. They also realized, however, that simple models with fewer decision variables were more playable and were initially more acceptable to players.

In a similar sense regarding game realism and student involvement Dukes and Walker (1976), after having evaluated six short games for realism, found a game’s subjectively-assessed value was a function of the following game features:

1. Accuracy -- the degree a game represents the model of reality that it purports to describe.
2. Plausibility -- the degree a game takes cognizance of the experience and capabilities of the players (Elder, 1973).
3. Relevance -- the degree a game relates to the concerns of the players.

In this case the game’s accuracy was only one of three factors leading to the perceived value of the gaming experience.

Boocock (1972) among many others (Hermann, 1967; Kibbee, 1967; Meier, Newell and Pazer, 1969; Mirth, 1972; Norris, 1986; Shubik, 1970) has addressed the various validity bases, which can exist for a business game. In her evaluation of intergenerational relations game three validities were examined: face validity, empirical validity, and theoretical validity. Of the three, face validity was found to be both an elusive game quality and a very important one as it was related to the game’s ability to capture the participants’ interest and involvement.

It could also be reasoned, however, that concerns about a game’s algorithmic validity are basically misplaced as any business game application is a patently false situation which in itself makes the experience invalid (Baldwin, 1969). Every teaching situation is administratively manipulated. Therefore, due to either the game’s simplification of reality, the isolated and protected nature of the learning situation, or from an “administration effect” brought about through behavioral cues emitted by the instructor, the external validity of a game-based teaching environment is suspect. In addition to simplified reality models, both Cohen and Rhenman (1961) and Greenlaw, Herron and Rawdon (1962) have noted the typical game uses fictitious factories, products, funds, resources and organizational responsibilities, a set of unique simulation roles easily obtained but relatively sparse information, and equal starting positions and lock-step decision rounds for all companies. Additionally, as a game increases in its complexity the need for instructor guidance increases, as naive players must be given help to keep them from floundering or engaging in “incorrect” playing strategies. Waggener (1981) found, however, that although students are aware of a game’s realism deficiencies and the administrator’s intrusions, most accepted the gaming experience as being both valid and worthwhile.

Decision Theory and Cognitive Psychology

The initial business gaming situation puts the typical game participant into a relatively ambiguous environment. This would be especially true for the top management games which are the most often employed game (Faria, 1987). These simulations create strategic decision making situations possessing imprecise, confusing and weak action cues, long creation and feedback time spans, and diverse functional relationships. Given this situation an individual’s cognitive structure has a great influence on what can be learned, and what mental restructuring of the game’s objective reality are being employed by the participant’s (McCaskey, 1976; Schweiger, 1980). When dealing with ambiguous decision making settings, it has been found that individuals simplify the situation (Tversky and
Hypothesis tested whether teams accepting a game’s realism will not be systematically related to the team’s economic performance.

H4 The detection of a program error will not be systematically related to the team’s economic performance.

**METHODOLOGY**

Students (n=100) from two undergraduate capstone business policy courses were randomly assigned to three-member firms playing The Business Management Laboratory (Jensen and Cherrington, 1983). Five industries possessing identical parameters were created except for the insertion of a glitch or program error in the advertising routine of three industries. This glitch caused advertising expenditures, a major demand creation component within the simulation, to have little demand effect on each firm’s primary product in its initial marketing area. A pre-game parallel test with and without the glitch found the error caused demand for the product to be 34.7% lower than normally expected in the primary marketing area and that extrapolated demand forecasts into the secondary marketing area would be underestimated by 44.7%.

After the simulation had ended participants responded to a two-part questionnaire. The first part employed a set of five-point Likert scales to evaluate each player’s perception of the realism possessed by a diverse ten item sample of the business game’s algorithms. To insure that the examined algorithms were representative and comprehensive, participants could volunteer up to three additional algorithms to the questionnaire; few additions were offered and no discernible response pattern existed within the nominations. Accordingly, the algorithms studied and listed in Table 1 are considered representative of all those that could have been examined in this study.

**TABLE 1**

**ALGORITHMIC FEATURES STUDIED**

1. The firm’s stock price as it relates to company earnings, stock offerings, and dividends.
2. The effect of the economic indicator on real growth in the simulated economy.
3. The effect of engineering studies on lowering manufacturing costs.
4. Each product’s degree of price elasticity.
5. Raw material “futures” being less expensive than “at market” raw materials.
7. The interaction of R&D budgets with new products or special features.
8. The effect sales representatives’ commissions have in stimulating product demand.
9. The automatic collateral feature associated with short term investments of differing levels of riskiness.
10. The effect of plant maintenance budgets on labor hour deterioration.

*Incorrect algorithm.
The questionnaire’s second part asked respondents to describe the features of any three previously listed algorithms. These descriptions (n = 380) were content analyzed according to the three perceptual accuracy conditions that could have existed: a perfect or completely accurate description of the features of the algorithm, the citation of an algorithmic feature that does not exist (a Type I error), or the statement of the nonexistence of an algorithmic feature that does exist (a Type II error).

**RESULTS**

An ANOVA was used to determine if a significant difference existed between the control and experimental groups regarding their ability to detect the advertising algorithm error. The procedure produced an F-ratio of 1.014 with a P-value of .467 indicating no significant difference between the two groups. H1 is therefore accepted. As shown in Table 2, the content analysis found that 60.78% of the perceptions of the incorrect algorithm were generally accurate while 19.61% of the descriptions committed a Type I error and an identical proportion committed a Type II error.

<table>
<thead>
<tr>
<th>Question</th>
<th>Perfect Perception (n=380)</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect algorithm</td>
<td>31</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>60.78%</td>
<td>19.61%</td>
<td>19.61%</td>
</tr>
<tr>
<td>Remaining questions</td>
<td>199</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>60.49%</td>
<td>27.36%</td>
<td>12.16%</td>
</tr>
<tr>
<td>All questions</td>
<td>230</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>60.26%</td>
<td>26.32%</td>
<td>13.16%</td>
</tr>
</tbody>
</table>

Chi-square 2.85, n.s.

A 60.49% agreement existed between the subjects’ perceptions of the remaining algorithms and each algorithm’s objective reality with 27.36% of the descriptions committed a Type I error and 12.16% of the descriptions committing a Type II error.

H2 stated that teams giving a higher realism score to the game’s algorithms would not be able to detect the error with any greater frequency than those teams with a lower realism score. Based on the ANOVA results presented in the previous section it is possible to conclude that regardless of the realism score, there was not a difference between the teams as to their error detection ability. To provide evidence regarding this question realism scores were regressed against the ability of each team to detect the error. The regression procedure resulted in R2 = .18 that indicates a very weak relationship between each team’s overall realism score and its error detection ability. Therefore was not rejected.

H3 stated a team’s realism score would not be related to its economic performance. To address this hypothesis the earnings of each team were regressed against their overall realism score. The regression procedure resulted in a non-significant F-ratio (P-value = .648) and an R2 approaching zero. Based on this result it can be concluded that no significant relationship existed between the team’s profitability and its realism score. H3 is therefore not rejected.

H4 stated that a team’s error detection ability would not be systematically related to the team’s economic performance. To test this hypothesis each team’s profits were regressed against its realism score. The regression resulted in an F-ratio of .300 (P-value .743) and a R2 again approaching zero. Based on these results it can be concluded that a non-significant relationship existed between these two variables and H4 is not rejected.

**DISCUSSION**

As found by the acceptance of all the acceptance of all the null hypotheses the presence of the glitch had no discernible effect on the player’s perceptions of the game’s realism, nor on each team’s economic performance. As shown in Table 3 all perceptual realism judgements were statistically equal and non-significantly different variation also existed within each algorithm examined. Given the non-detection of the program error, all players felt the simulation was fairly realistic.

Although it was theorized that significant perceptual differences should exist, these differences were not found. A number of further studies are warranted to more fully test the hypotheses stated. In the present study only one algorithm was examined and that algorithm dealt with a very ambiguous, non-deterministic function within the firm. As shown in the Wolfe and Teach (1987) game review, the BML’s overall demand equation is a complex, multiplicative function embracing the absolute and relative number of sales reps and their compensation incentives, the markets absolute scale, price levels and price changes, and product style features and product quality in addition to the advertising component was masked, or lost in the cross currents of other demand creation features found in the game’s marketing routines.

It is also possible that the market feedback nature of the algorithm’s effect also made error detection very difficult. Given the mixed signals players obtain from any market-driven business game, it is difficult to determine the effectiveness of various company actions, especially since the actions by one team often spur countervailing reactions from other teams. It is possible the introduction of algorithmic errors into the simulation’s more deterministic routines, such as in its depreciation expense schedule, or its administration expenses, or its production function/raw materials ordering function, would make the error more easily detected.

One additional game related element could also explain the results found in this study. It has been determined the BML is a relatively complex game containing 1,118 executable statements in its FORTRAN program. As shown in the table reproduced here from Butler, Pray and Strang (1979) p. 485 the crippling of one or two algorithms in the Simple game might be more easily seen or felt than if the same number of algorithms were crippled in the ADSIM game or the Complex game.

**CONCLUSION**

A deliberate glitch or program error was inserted into the demand creation function of a relatively complex business game as a partial test of a simulation’s need for algorithmic accuracy. It was theorized that it would be difficult to detect the error and that the presence of the error would not materially affect the quality of play as decision makers possess their own sense of reality and the simulations program is only one aspect regarding the determination of a game’s realism. Players were unable to detect the error and the error’s presence had no impact on the quality of
TABLE 3
QUESTIONNAIRE RESPONSES

<table>
<thead>
<tr>
<th>Algorithmic Function</th>
<th>Mean Response</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The firm’s stock price as it relates to company earnings, stock offerings, and dividends.</td>
<td>3.76</td>
<td>.94</td>
</tr>
<tr>
<td>2. The effect of the economic indicator on real growth in the simulated economy.</td>
<td></td>
<td>.97</td>
</tr>
<tr>
<td>3. The effect of engineering studies on lowering manufacturing costs.</td>
<td>3.63</td>
<td>1.11</td>
</tr>
<tr>
<td>4. Each product’s degree of price elasticity.</td>
<td>3.56</td>
<td>1.09</td>
</tr>
<tr>
<td>5. Raw material “futures” being less expensive than “at market” raw materials.</td>
<td>3.82</td>
<td>1.07</td>
</tr>
<tr>
<td>6. The effect of advertising expenditures on product demand.</td>
<td>3.46</td>
<td>1.17</td>
</tr>
<tr>
<td>7. The interaction of R&amp;D budgets with new products or special features.</td>
<td>2.84</td>
<td>1.37</td>
</tr>
<tr>
<td>8. The effect sales representatives' commissions have in stimulating product demand.</td>
<td>3.77</td>
<td>1.01</td>
</tr>
<tr>
<td>9. The automatic collateral feature associated with short term investments of differing levels of riskiness.</td>
<td>3.21</td>
<td>.85</td>
</tr>
<tr>
<td>10. The effect of plant maintenance budgets on labor hour deterioration.</td>
<td>3.80</td>
<td>.97</td>
</tr>
</tbody>
</table>

Mean 3.53

| Very Realistic | 5.0 |
| Very Unrealistic | 1.0 |

TABLE 4
PROGRAM COMPLEXITY BY BUSINESS GAMES

<table>
<thead>
<tr>
<th>Statements</th>
<th>Simple</th>
<th>Intermediate</th>
<th>ADSIM</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computations</td>
<td>44</td>
<td>137</td>
<td>462</td>
<td>578</td>
</tr>
<tr>
<td>IF</td>
<td>13</td>
<td>35</td>
<td>72</td>
<td>153</td>
</tr>
<tr>
<td>WRITE</td>
<td>25</td>
<td>22</td>
<td>83</td>
<td>136</td>
</tr>
<tr>
<td>DO LOOP</td>
<td>5</td>
<td>12</td>
<td>35</td>
<td>111</td>
</tr>
<tr>
<td>GO TO</td>
<td>4</td>
<td>28</td>
<td>29</td>
<td>91</td>
</tr>
<tr>
<td>READ</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>49</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>245</td>
<td>693</td>
<td>1118</td>
</tr>
</tbody>
</table>

*From Butler, Pray & Strang (1979, p. 485)

their economic performance. To more completely determine the need for a simulation to possess algorithm accuracy, the work begun in this study should be expanded to include the examination of (1) more error-ridden algorithms in complex games, (2) a smaller number of incorrect algorithms in simple games, and (3) errors in a simulation’s deterministic functional areas. The results of this study also highlight the need for the instructor to help players understand their personal reality sense and to rationalize the reality presented to them by the simulation.

REFERENCES


Developments in Business Simulation & Experiential Exercises, Volume 16, 1989


Balance of References will be supplied by the senior author: Joseph A. Wolfe, The University of Tulsa, 600 South College, Tulsa, Oklahoma 74104