

Developments in Business Simulation and Experiential Learning, Volume 28, 2001
AN INITIAL VALIDITY INVESTIGATION OF A TEST
ASSESSING TOTAL ENTERPRISE SIMULATION LEARNING

Jerry Gosen , University of Wisconsin -Whitewater

John Washbush, University of Wisconsin -Whitewater

ABSTRACT

This paper is the third in a series dealing with the construction of a test bank of items designed to assess the degree to which learning takes place from playing a total enterprise simulation. It provides data as to whether the test central to this research is valid. In this study, relationships between results on this study's learning test and two criterion variables --self-report of learning and forecasting accuracy--were examined. The results show at best a cloudy picture with respect to the test's validity, partially because of a low number of subjects (N=23). On one hand, the results reveal a negative relationship between learning scores on the test and forecasting accuracy. On the other, they show marginally significantly greater learning scores for those who said on a self-report measure that they learned the game's complexity and financial analysis.

INTRODUCTION

This study is part of a long-term project to develop instrumentation to assess learning from a total enterprise simulation. The project was proposed in the context of criticism of the simulation field for not defining or properly measuring the learning that takes place from simulation play. Among the critics were Anderson and Lawton (1997a), Gentry et al. (1998), and Thavikulwat et al. (1998).

In earlier phases, we developed a test bank of 122 multiple-choice and short essay items (Gosen et al., 1999) and gathered some initial data (Gosen, Washbush & Scott, 2000) on two instruments from the bank. The purpose of the present study was to initiate an investigation of the validity of a third (hopefully improved) instrument. The test itself consists of 38 of the test bank's 122 items.

Validity

Thavikulwat, et al. (1998) have proposed these standards for evidence of validity for assessment instruments: 1) show evidence of reliability, 2) be able to discriminate between individuals with different levels or types of learning, 3) show convergence with other instruments attempting to measure the same constructs, and 4) yield normative scores for different populations. The validity of a test score, according to McDonald (1999), is the extent to which

it measures an attribute of the respondents that the test is employed to measure in the population for which the test is used. Alternately, a test is valid if it measures what it purports to measure.

Given Thavikulwat categories, we've shown evidence of reliability in previous studies (Gosen et al., 1999; Gosen, Washbush & Scott, 2000), and we've argued elsewhere (Washbush & Gosen, under review) that it would be easy to attain normative scores for different populations. The present study focuses on convergent validity. Convergent validity according to McDonald (1999) is when scores on a test are highly correlated with scores (often called criterion measures) on other measures reflective of the same construct.

This investigation attempted to focus on two such criterion measures. The first is forecasting accuracy. This variable has been proposed by Teach (1989, pg. 103) as...(the indicator)...of proficiency with which managers and (student simulation participants) execute a critical management process which is highly associated with a firm's success. He argues further (Teach, 1990, pg. 21)...that forecasting is a learned skill and that one would expect students to get better with practice. Anderson and Lawton (1988, pg. 242) contend that forecasting accuracy reflects a team's ability to translate its decisions into simulation outputs.

The second criterion is self-reported learning. This is a subjective measure, and its use has been criticized (Gentry et al, 1998), but it makes sense that how much one learns ought to be consistent with how much she thinks she learns.

Background

We contend that that this project is the first attempt to create instrumentation to assess simulation learning from specific objectives emerging from the simulation itself. There have been attempts to measure simulation learning, but the measuring devices have often been indirect, including, for example, course grades (Comer and Nichols, 1996) and course exams (Raia, 1966 and Wellington and Faria, 1991). There have been measures that are more direct but stem from very general learning objectives such as attaining quantitative skills (Faria and Whiteley, 1990 and Whiteley and Faria, 1989), company self-concept development, (Pearce, 1978-90), and goal-setting abilities (Wheatley, Horneday and Hunt, 1988). One study in which learning measures have emerged from specific learning goals was

Developments in Business Simulation and Experiential Learning, Volume 28, 2001

performed by Wolfe (1976). His focus was on the effects of game participation on learning strategic management and organizational goal setting. His more specific objectives included 'administer a preconceived strategy' and 'create the components of a business policy system.'

After reviewing the above studies it appears that in only one study (Wolfe, 1976) were specific objectives used to guide the development of an instrument measuring simulation related learning, and in none were measurement devices developed from specific objectives emerging from the simulation itself. The present research was designed to fill the void. For the test bank central to this research, the items created were developed from specific objectives emerging from the simulation (Gosen et al, 1999). The long-term result of this effort is intended to be a test bank of usable items, the objectives from which they emerge, and reliability and discrimination statistics. The intention is also to create simulation-learning-related scales and validity statistics for each scale.

METHOD

Subjects and Procedure

Twenty-three students taking the capstone policy course at the University of Wisconsin-Whitewater during the summer of 2000 participated. They played seven quarters of MICROMATIC (Scott et al., 1992), preceded by a practice round. Prior to the practice round, students were administered version 3 of the learning test central to this study as a pre-test. They were also asked in an open-ended question what they expected to learn from the simulation. After the game ended, students again completed version 3 of the test as a post-test and also responded to an open-ended question about what they learned from playing. With each decision, they were required to forecast company sales in units. Game performance and the post-test score were each worth 12 ½ % of the students' course grades. Five percent of the course grade was based on peer ratings of team contribution.

Variable Measurement

Forecasting Accuracy was the total of the differences

between predicted sales in units and actual sales in units for all quarters.

Learning for each participant was defined as post-test percent score minus pre-test percent score. We used a common scoring key to ensure uniformity of measurement. Kuder-Richardson reliability coefficients were .707 for the pre-test and .724 for the post test.

Performance in the simulation was measured at the end of play using the game's scoring procedure and was based on net income (40%), return on assets (30%), and return on sales (30%).

Self-Report of Learning was measured with the use of the open ended question, "What did you learn by playing the simulation?" The responses were content analyzed into 10 categories:

- game complexity
- general cause and effect
- specific cause and effect
- keeping production, warehousing, and marketing in balance
- forecasting
- principles applicable to business in general strategy
- dealing with mistakes
- planning
- financial analysis

RESULTS

Table 1 contains these results pertinent to the use of forecasting accuracy as a criterion variable. Learning scores correlated negatively and significantly ($r = -.44$; p less than .05) with forecasting accuracy for all forecasting attempts. Those who learned the most had the *largest* discrepancy between their sales forecast and actual sales. Learning also correlated negatively to a greater degree with forecasting accuracy for earlier quarters ($r = -.48$; p less than .05) than for later quarters ($r = -.19$). Additionally, forecasting accuracy correlated significantly and positively ($r = .87$; p less than .001) with the profits-dominated simulation performance measure in this study, a result consistent with results found by Teach (1989). Finally, learning and performance correlated negatively and significantly ($r = .50$; p less than .05).

Table 1
Correlations Between Learning, Performance
and Forecasting Accuracy

	Learning Score	Post-test	Performance	First 3 Forecasts	Last 4 Forecasts	All Forecasts
Learning Score	1	-.22	.50	-.48	-.19	-.44
Post-test		1	-.03	-.38	-.20	-.49
Performance			1	.33	.78	.87
First 3 Forecasts*				1	.11	.49
Last 4 Forecasts*					1	.81
All Forecasts*						1

Table 2
Learning Score as a Function of Self-Report Categories

	Not Stated as Learned			Stated as Learned			t	p
	#	Mean Learning Score	Variance	#	Mean Learning Score	Variance		
Game Complexity	14	.034	.003	9	.084	.009	1.56	.07
General Cause & Effect	9	.052	.006	14	.069	.006	0.46	.33
Specific Cause & Effect	15	.073	.008	8	.042	.004	0.98	.17
Balance	19	.070	.007	4	.030	.003	0.94	.19
Forecasting	12	.056	.006	11	.070	.008	0.38	.35
General Business	12	.064	.006	11	.061	.009	0.08	.47
Strategy	16	.062	.006	9	.066	.010	0.11	.46
Mistakes	13	.067	.006	10	.060	.008	0.27	.39
Planning	19	.057	.007	4	.084	.008	0.55	.31
Financial Analysis	15	.046	.008	8	.093	.004	1.49	.08

Table 2 contains the results relevant to the use of the self-report of learning as a criterion. As noted above, content analysis of the response to the question, “what did you learn by playing the simulation?” revealed ten categories. This table displays t-tests with learning scores as the dependent variable and whether or not one reported a certain category of learning as the independent variable. This table shows that those who reported that they learned the complexity of the game had almost significantly higher learning scores than those that did not ($t=1.56$; $p=.07$) and those who reported that they learned financial analysis had almost significantly higher learning scores than those that did not.

DISCUSSION

The results of this study do not support claims for instrument validity. Although not established as a valid criterion for the learning construct, there are claims (Teach, 1990) that forecasting improves with practice and varies as participants learn the game. That a learning score on a test does not vary positively with the accuracy of forecasts suggests that either the test, the forecasting variable, or both are not valid representations of the learning construct.

There are many possible reasons for a negative relationship between the learning score and forecasting accuracy. First, because of the short (3-week) academic session, time permitted only 7 decisions, and it is possible that one

Developments in Business Simulation and Experiential Learning, Volume 28, 2001

or more of the measures of the variables in this study (forecasting, profits, or learning) are not reliable indicators over such a small time span. It is possible that learning does not take place in a simulation until after many decisions, thus learning after seven decisions would not be a reliable gauge. This argument has been offered explicitly by Teach (in conversation) who has argued that it may take twenty or more decision for learning to take place, and, once it does, it may influence performance, and a positive learning-performance relationship may emerge. Second, forecasting may be more closely related to performance than learning. Evidence supports that notion. For example, Anderson & Lawton (1992b), Thorngate & Carroll (1987), Washbush & Gosen (under review), and Wellington & Faria (1991) have all found a *lack* of a statistical relationship between learning and performance. Perhaps forecasting measures the same thing as performance, while the learning score in this study reflects something else.

We cannot make strong claims for the validity of this study's instrument on the basis of two almost significant relationships between learning score and two self-reports of learning variables. First, these results were significant at just less than the .10 level, hardly a strong justification for a conclusion. Second, the criterion variable is a self-report measure, a type of measure often criticized (Gentry et al., 1998).

CONCLUSION

This study suggests a model for the validity analysis of a test measuring learning acquired from participating in a simulation. Unfortunately, small numbers and the constraints of a very brief and rushed academic term raise important questions about the results obtained. This model is appropriate for continuing investigation in normal, semester-long academic settings and over larger sample groups. In the present study, players only forecasted unit sales. In contrast, Teach's players (1989) forecasted net income and cash balance in addition to sales.

REFERENCES

- Anderson, P.H. & Lawton, L. (1997a). Demonstrating the learning effectiveness of simulations: where we are and where we need to go. *Developments in Business Simulation & Experiential Exercises*, 24, 68-73.
- Anderson, P.H. & Lawton, L. (1997b). Designing instruments for assessing the effectiveness of simulations. *Developments in Business Simulation & Experiential Exercises*, 24, 300.
- 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.
- 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.
- 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.
- Gosen, J., Washbush, J., Patz A., Scott T.W., Wolfe, J., & Cotter, D. (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.
- McDonald, R.P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates, London
- Pearce, J.A.II (1978-1979). Developing business policy skills: A report on alternatives. *Journal of Educational Technology Systems*, 7, 361-371.
- Raia, A. P. (1966). A study of the educational value of management games. *Journal of Business*, 39, 339-352.
- Scott, T.W., Strickland, A.J., Hofmeister, D.L. & Thompson, M.D. (1992). *Micromatic: A Management Simulation*. Boston: Houghton Mifflin.
- Teach, R., (1989). Using forecasting accuracy as a measure of success in business simulations. *Developments in Business Simulation & Experiential Exercises*, 16, 103-107.
- Teach, R., (1990). Profits: the false profit of business gaming. *Simulation & Gaming: An International Journal*, 21, 1, 12-26.
- Thavikulwat, P., Anderson, P.H., Cannon, H., & Malik, D. (1998). Games as instruments of assessment: A framework for evaluation. *Developments in Business Simulation & Experiential Exercises*, 25, 31-37.
- Washbush, J. & Gosen, J. (under editorial review). Learning in total enterprise simulations. *Simulation & Gaming: An International Journal*.
- 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.
- Wheatley, W.J., Hornaday, R.W. & Hunt, T.G. (1988). Developing strategic management goal setting skills. *Simulation & Games*, 19, 173-185.
- 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.