

Developments in Business Simulation & Experiential Exercises, Volume 9, 1982

PROBLEMS ASSOCIATED WITH THE ASSESSMENT OF EXPERIENTIAL LEARNING USING THE MULTIPLE CHOICE TEST

Ronald O. Klein, Columbus College

ABSTRACT

The multiple choice test medium was used in a before- and-after test format to assess the degree of learning from a business simulation game. During the test quarter, an experimental group participated in a multinational game and heard three 2-hour lectures on international business topics. The control group played a domestic business game and heard the same lectures. Consequently, the test data attempt to compare the amount of international business knowledge acquired under these two test conditions: simulation plus lecture and lecture only.

The before-and-after difference between groups was not statistically significant as measured by the number of correct responses, yet side by side open-ended questions recorded substantial between-group differences. To explain this apparent weakness in the multiple choice test, an in-depth analysis of the test's validity was undertaken. The analysis, using multivariate techniques, shows strong evidence of learning in spite of the lack of statistical significance. This leads to doubt concerning the suitability of the highly structured multiple choice format for experiential learning assessment and, further, indicated that care should be taken in hypothesis testing using parametric statistics.

INTRODUCTION

A review of the literature shows that there is a lack of consensus concerning the validity of the business simulation game as an effective learning medium. Experimentation continues to produce mixed results. Part of the problem of measuring "learning" may be in the selection of test instruments. The multiple choice (MC) test, which is almost universally accepted as a valid measurement tool in the text/lecture teaching format, may be ill-suited to the relatively unstructured learning environment of the business simulation game. This paper examines this premise by subjecting a set of MC test results to rigorous analysis.

THE EXPERIMENTAL SETTING

Seventy business majors, at the senior policy level, were divided into two groups; one of which played a domestic policy game, the other played an international business policy game. Neither group had any significant prior introduction to either international business or simulation games. The students were given a pre-test and a post-test multiple-choice examination consisting of sixty-eight international business questions; they were also asked to answer short, essay questions.

The Essay Test

The essay test results showed students in the different test

groups were sensitized to critical issues related to their respective business environments, either domestic or international. In response to general essay questions on planning, for instance, the multinational competitors expressed concern for inflation and expropriation; their domestic counterparts were more concerned with raw material shortages and labor costs. In responding to specific international questions, the multinational group provided decidedly superior answers to those questions which were related to the gaming experience. Keyword counts were used to assess test performance to reduce subjectivity in grading. This test was given at mid-term prior to lectures on international business. All observed differences can be attributed to participation in the simulation experiences and the interest generated during these experiences.

Table 1 shows no significant difference between groups in response to the questions; "How will social responsibility toward customers change in the near future?" As a manager what specific problems do you think will arise?" Neither simulation game addressed this subject area in terms of its required decisions. No differences were expected, none were observed.

TABLE 1
SOCIAL RESPONSIBILITY
KEY WORD COUNT

	<u>MNG</u>	<u>TEMPOMATIC</u>
Environment protection	11	14
Product safety	13	12
Product quality	6	8
Consumer needs	8	6
Community responsibility	2	2
Worker benefits	1	4
Energy conservation	2	4
Other resource management	2	4
Truth in advertising	0	1
Total number of key words	45	55
Number of test subjects	13	14
Average number of key words	3.5	3.9

Table 2 focuses on financial risk, a problem directly addressed by both games. The question was: "Every company is exposed to a variety of risks: to what risks are your company's assets exposed?"

Using a Chi Square contingency table analysis, the difference is significant at the .001 level of confidence. Note particularly the differences with reference to foreign exchange losses and expropriation. This pattern of results was consistent throughout the short answer, essay test. There were significant differences on all six game topic related questions.

Developments in Business Simulation & Experiential Exercises, Volume 9, 1982

TABLE 2
BUSINESS RISK QUESTION
KEYWORD COUNT

	<u>MNG</u>	<u>TEMPOMATIC</u>
Expropriation	36	1
Inflation	13	9
Competition	10	8
Regulations	2	2
Obsolescence	2	7
Interest rate	8	4
Exchange loss	18	0
Theft	2	5
Consumer preference	5	6
Shortages	2	5
Natural disasters	0	4
Total number of key words	98	51
Number of test subjects	34	33
Average number of key words	2.9	1.5

The Multiple Choice Test

A sixty-eight question MC test focusing entirely on international business concepts and issues was given to both groups before and after the simulations were played [1]. In addition, the games were supplemented by three 1-hour international business lectures during this quarter long study. Since both groups were present during the lectures, any differences in test results should be due to the gaming experience. The results are summarized in Table 3.

TABLE 3
QUESTIONNAIRE DATA PRE- VERSUS POST-TEST
(NUMBER OF CORRECT RESPONSES)

	<u>Pre- test</u>	<u>Post- test</u>	<u>Sam- ple</u>	<u>Size of Improve- ment</u>	<u>t- Score</u>
Experi- mental	31.9	36.1	34	4.2	
Control	31.8	34.5	33	2.7	
				Critical	t = 1.3 .05

This test fails to support the results from the essay tests. These differing findings prompted a more detailed analysis of the MC test instrument.

TEST INSTRUMENT VALIDITY

Initially, each question was subjected to a one-way ANOVA to examine its ability to discriminate between test groups on the pre-and post-tests. On the pretest, 36% of the questions yield an F value greater than 1; on the post-test, 57% fall into this category. This indicates that there was a greater percentage of correct responses by the experimental group on the post-test as "measured" by the F-ratio. This difference seems much more substantial than the differences indicated by the raw scores in Table 3.

One would also hypothesize that the response variances would be smaller. Students would tend to guess less, narrowing their range of responses. The average standard deviations on the pre- and post-tests were .368 and .319, respectively. This is an obvious difference but of lesser magnitude perhaps than one might expect. More on this later.

Next, a discriminant function was developed to assess the predictive ability of these variables. The coefficient of determination increased from .695 to .777 from the pre-test to the post-test. The percent of correctly classified cases increased from 70.3 to 91.7.

Arguments can be made against any of these findings individually, but it is difficult to refute the findings as a whole. By the end of the quarter, variances narrowed, the number of discriminating questions increased, and a discriminant equation improved. In other words the underlying structure of the response patterns consistently improved, moving in the direction one would expect if learning had taken place.

Communality Analysis

Classical factor analysis is based on the assumption that a specific set of variables has some underlying relationship. Since almost all of the data questions are related to international business we can safely assume that such a relationship exists on the macro level, i.e., all questions are related to a single factor. The strength of this relationship can be estimated using a measure of shared determinants. The part of the variable influenced by shared determinants is called "common," the part influenced by idiosyncratic determinants is called "unique." If the communality among the variables increases from the pretest to the post-test, there is reason to believe that learning has taken place with reference to international business. The size of the communality index is useful for assessing how much of the variance of a particular variable is accounted for by all factors together. Large communalities indicate that a large amount of the variance is accounted for by the factor structure [2].

Table 4 below shows the estimated communalities in a pre-test versus post-test comparison. The t-test statistic at the foot of the table shows that the difference is significant. This statistic appears to be much more sensitive to between-test differences than the previously mentioned distribution variances.

TABLE 4
MEAN COMMUNALITIES

	<u>X</u>	<u>SD</u>
Pre-test	.5449	.0927
Post-test	.7679	.0949

Calculated t = 1.3
Critical t = 1.3
.05

These results indicate that the response patterns for the groups as a whole changed significantly from the pre-test to the post-test with reference to international business key questions and issues.

The foregoing analysis indicated that response patterns changed during the test period which in turn strongly supports the hypothesis that learning has taken place. Since the simulation played an important role with reference to the international business material, there is a sound basis for suspecting a cause and effect relationship between learning and simulation.

Developments in Business Simulation & Experiential Exercises, Volume 9, 1982

The Factor Structure

The next step is to move past the “general” measures of the factor analysis and examine the factor structure. Differences in the underlying factors between groups at the end of the study (post-test) would not only be indicative of learning but may be useful in isolating what is learned in the experiential situation. There were four findings from this portion of the analysis.

First, the MNG group yielded factors easier to define. Questions on finance and marketing, for instance, tended to appear under a single factor. The factors for the TEMPOMATIC group were more difficult to name because of the variety of subjects addressed by questions loading on a particular factor. Second, the test questions which loaded higher on factors were the short, definitional questions, such as: “Eurodollars are,” “A joint venture is.” A respondent generally knew the answer or not. The wordy, more general questions were poorer between group discriminators, e.g., “The forward market is used by businessmen to.” Third, of the low discriminating questions, more than 80% had as a response alternative “all of the above.” Less than half of the higher discriminators had this phrase as one of its answers. Finally, the questions related to game decisions appeared to be much better discriminators than questions less related to game decisions.

In summary, questions that differentiate best between groups are unambiguous, definitional, game (decision) related questions. There is no statistical proof that learning took place based on MC test performance but there is reasonable doubt concerning the validity of the test as used in this study.

CONCLUSIONS

There are some serious technical problems associated with the MC measures used in the analysis above. First let us review the primary measure of central tendency: the arithmetic mean. The response means are meaningless since they are dependent on which of the multiple choice answers is the correct one. The variances around these means are more meaningful since they are indicative of the differences of opinion of the group concerning a particular answer. The variance is not a perfect measure, however. For example, suppose the correct response to a question is number one, but only half of the group selects number one and half selects number five. In this extreme case, the mean would be 2.5 and the variance, a measure of dispersion, would be very large even though the group limited itself to two of the five choices. In fact, all other combinations of choices would result in a smaller variance even though more choices are involved. This is the extreme case. In general, the variance will correlate positively with the range of choices, but statistically perilous problems are present.

Much of our research literature employs the ubiquitous MC test in various leaning measurement studies, and the results have been mixed with reference to the effectiveness of the simulation as a learning tool. In this study, mean scores were poor measures of group differences. The resulting variances were seriously affected by the nominal or categorical nature of the data. Subsequently, more detailed analysis indicates substantial between group differences. Consequently, we can conclude that there is an important technical problem associated with the use of parametric statistics. The simulation appears to be effective, but we must construct our test instruments more carefully. The typical “instructor’s manual” MC test designed for lecture and text learning is less than appropriate for the measurement experiential learning.

Finally, when the carefully constructed MC test is used to measure experiential learning, appropriate statistical

techniques should be applied when testing research hypotheses.

REFERENCES

- [1] Daniels, John D; Ogram, Ernest W. Jr.; and Radebaugh, Lee H., International Business: Environments and Operations (Reading, Mass: Addison Wesley Publishing Co., 1976).
- [2] Hair, Joseph F., Jr., et al, Multivariate Data Analysis (Tulsa, Okla., Petroleum Publishing Co., 1979).