Developments in Business Simulation and Experiential Learning, Volume 26, 1999

A STUDY OF THE ETS GENERAL FIELD TEST AS AN AACSB ASSESSMENT TOOL AND THE IMPACT OF EXPERIENTIAL EXERCISES AND SIMULATION ON LEARNING

Barbara Howard, SUNY at Geneseo
Peter Markulis, SUNY at Geneseo
Daniel Strang, SUNY at Geneseo

ABSTRACT

This paper explores the use of the ETS exam as an assessment tool and as a possible indicator of the effectiveness of experiential learning techniques. The authors use a sample of students that completed the ETS general field study test to analyze the relationship between several demographic factors and the results on the test. No relationship was found between the mean score on the test and classroom exposure to case study, computer simulation, or experiential exercise techniques. The study, however, was severely limited by the small sample size. The authors conclude that although the design of the study is valid, the results are not.

INTRODUCTION

Experiential exercises and simulations have long been touted as popular teaching mechanisms, and classroom usage continues to increase. However, considerable debate continues over the contribution to learning that these methods provide. At the same time, outcomes assessment has become a widely accepted practice, both as a means of self-improvement and as an integral part of many accreditation processes, including AACSB, and the International Association for Management Education. The combination of these two factors led the authors to review one of the more commonly used assessment tools, the ETS General Field Test. It is widely held by many, that tests such as the ETS General Field Test represent a valid and reliable method for measuring student learning in the management curriculum, including the learning from experiential exercises and/or simulations.

LITERATURE REVIEW

Learning Effectiveness

Many of us contend that there is a benefit associated with the incorporation of experiential exercises in our classes; a benefit that outweighs the sometimes high administrative costs. Efforts to "prove our point" i.e., that experiential learning techniques actually provide a value added component to the learning process, have been ongoing and numerous. Unfortunately, copious research has been completed and analyzed without hard evidence that experiential exercises and simulations improve the learning process (Anderson & Lawton, 1997). Greenlaw and Wyman (1973) started the debate by questioning the effectiveness of simulations as a classroom technique. Whitely and Faria (1989) found no impact on exam performance as a result of simulation participation. Wellington and Faria (1991) found no relationship between recency of play and performance level on exam performance. Nonetheless, Keys and Wolfe (1990) argue that simulations are an effective pedagogical method while several others, including McKenna, (1991); Washbush and Gosenpud, (1991); Leonard and Leonard, (1995) have found that participant attitudes and reactions to simulations have been very positive. Such results clearly favor a high disposition toward learning, if not to the results of learning *ipso facto*.

In recent years, a number of researchers have put forth new methods of operationalizing the learning criteria. Sackson (1992) presented cluster analysis as a means of measuring both game and student performance. Biggs, Miles and Schubert (1993) developed an instrument to

measure *perceptions* of effectiveness; and Burns (1993) proposed a grid comparing the effectiveness and efficiency of the teams as measured by market share and profitability respectively.

While the research to date has not been unanimously supportive of the effectiveness of simulations and experiential exercises as learning tools, there are a number of issues that could be addressed which would improve the validity of the findings (Malik and Howard, 1996). Herz and Merz (1998) contend that "learning as a research objective is much too broad and too vague to be controlled for the evaluation of a simulation/game." It appears that a clear hypothesis about the process of learning is necessary in order to design a study that will provide meaningful conclusions. This is obviously easier said than done. Thus, it seems that we are destined to continue to test, analyze, and debate the effectiveness of simulations and experiential exercises as learning tools. Given the difficulty of designing an "uncontestable" study, many researchers, as well as practitioners, resort to a second front, that of the surrogate measure. One such popular surrogate measure is an outcomes test. Such tests ignore the complexities of input and process variables and simply attempt to measure what the curriculum was supposed to have imparted to the student, the subject matter or some reasonable facsimile thereof.

Outcomes Assessment.

Historically, the quality of an educational institution was measured by evaluating the inputs or resources. These input measures included standards for teacher/student ratios, faculty composition, curriculum, library resources, student quality (standardized test scores, GPA, class rank), and faculty development activities. Little or no attention was paid to the quality of the outputs generated by the institution. The basic premise was that "quality in meant quality out." Today, however, this premise is no longer the norm, indeed it is being attacked by a variety of educational stakeholders. Parents, students, alumni, employers, and taxpayers in general have demanded more accountability from educational institutions at every level. This demand for accountability was a byproduct of the concern over several other issues, including dwindling resources, tuition increases that exceeded inflation, critical reports from both administrators and the general press, and a general dissatisfaction with America's educational system.

The groups that accredit educational institutions have reacted by creating task forces to determine how to allay the fears of a concerned public. One response was to develop outcome assessment policies and tests that would provide the stakeholders some degree of confidence in the quality of their institutions. While the major accrediting bodies recognize that institutions have diverse objectives and programs, with diverse outcome assessment programs, they expect a formal and regular process to be in place. Without prescribing an exact standard assessment model for all, they have identified a conceptual framework to follow. The key elements or steps include:

1. state the mission or purpose
2. formulate goals and objectives at both the institutional and program level that are consistent with the mission
3. develop and implement a set of procedures that can be used to evaluate the attainment of the goals and objectives
4. use the results of the assessment to improve
5. develop procedures to examine the extent to which the outcome measures effectively test the goals and objectives (Debrecency, 1997)

According to the accrediting associations, the role of an effective outcome assessment program should be a formative or improvement-oriented one rather than a summative or comparison-oriented one. The purpose should

be to facilitate improvement within an organization, not to generate quantitative measures by which to rank colleges. The diversity of the institutions and the danger of comparing outcomes across institutions led the AACSB to proclaim "No single set of measures suits all schools. Rather, the distinction of individual programs should be matched by distinctive sets of measures used to evaluate effectiveness. (AACSB Task Force on Outcome Measurement, 1989, p2) While the appreciation for diversity and therefore the potential for a variety of outcome assessment programs is evident in the AACSB statement, this does not mean that any program will be deemed acceptable. The accrediting associations want the individual institutions to develop programs that will work for them given their specific mission. Incorporating outcome measures into the planning process is part of continuous improvement. According to Milton Blood, director of accreditation for AACSB, "It isn't enough just to have outcome assessments. They need to be used in strategic planning" (Newsline, 1997)

A discussion of input measurements typically meant a finite set of standards that was applied to every institution however, the set of outcome measures is infinite. "The number and nature of the outcome measures developed and used by an institution is constrained only by its creativity and budget." (Debrecency, 1997) Accrediting bodies often list the type of outcome measures that an institution would find useful. Some of the more common ones mentioned are:

- Student surveys
- Alumni surveys
- Employer surveys
- Retention rates
- Graduation rates
- Employment placement rates
- Graduate school placement rates
- Senior theses
- Comprehensive exam scores
- Standardized achievement test scores
- External student honors/awards

It is not unusual for an institution to use multiple measures in their assessment process. Depending on the institution's mission, a multi-faceted approach may be the only way to assess whether there has been value-added.

RESEARCH DESIGN

As the authors' business school struggled to develop appropriate outcome assessment instruments and processes, the ETS General Field Test came under consideration. After considerable debate over several issues associated with the test, and finally agreeing (somewhat reluctantly) to the assumption that the ETS test is a valid indicator of achievement, the school decided to use this test as one piece of its outcome assessment. The method of administration also provoked lengthy discussions concerning cross-sectional reliability. The final result was that the school asked seniors to voluntarily prepare and complete a "standardized" test early in their last semester. The cost of the ETS test prompted the school to take a two-prong approach to this test. Thirty ETS tests would be purchased and a "home-made" test would be given to the remainder of the students. The "home-made" exam would be developed in-house by the faculty as a whole. Each faculty member submitted multiple choice questions that covered what he or she considered the pertinent topics of their particular discipline. A test was then compiled which resembled the ETS test with respect to the subjects covered, type of questions, and length of exam. The use of the "home-made" test would minimize the expense of administration (thus preserving departmental funds for faculty travel), and provide the opportunity for longitudinal testing using the "home-made" exam. A strong argument was made by some faculty that the developers of the curriculum should also be the developers of the assessment instrument. Thus, the "home-made" exam would be a better measure of what we, as a faculty, set forth as the objectives of the learning our students were supposed to accomplish.

The Management and Accounting majors were all asked to sit for the exam with the intention of assigning the exams randomly after assuring that all thirty ETS exams would be used. Unfortunately, (although not unexpectedly) only thirty students (approximately 20% of the total population) volunteered to sit for the exam. Therefore only the ETS test was utilized this year.

The authors decided to use this serendipitous circumstance to see if the achievement test results were impacted by the pedagogical methods to which the students were exposed. If the test is a valid measure of "learning" it seemed natural to see if the method of learning resulted in a higher (or lower) score. A questionnaire, designed to determine and control for several demographic and experience variables, was completed by the students just prior to the exam. Data regarding the student's major, GPA, exposure to various experiential learning methods, and transfer student status were collected. The authors hypothesized that some of these demographic factors would have an impact on the test results.

RESULTS

Thirty students were administered the ETS examination, as well as a brief demographic and background questionnaire. The background/ demographic questionnaire revealed the following:

- 24 students reported that they had taken a course (courses) that included the play of a computerized simulation
- 28 students reported that they had taken a course (courses) that included the use of an experiential exercise
- all 30 students reported that they had taken a course (courses) that included the use of case studies
- 14 students reported that they had completed an internship
- 13 of the students had transferred to the authors' institution from another institution
- 20 of the students were majoring in accounting and 10 were majoring in management

The authors had hoped to relate performance on the ETS to various background and demographic factors, however, with only 30 students included in the sample before it was sub-divided based on these factors, the analysis was limited by the minimal degrees of freedom. For example, in the instance of prior exposure to case studies, all 30 students reported that they had had this exposure. Thus, the authors were not able to separate out this factor as a treatment variable. Similarly, although not as pronounced, difficulties occurred for simulations and experiential exercises. Stated simply, with a maximum sample size of 30, too great a percentage of the students responded positively for several of the demographic/background factors and too few responded negatively, to permit meaningful analysis of the impact of these factors. As a result, a bi-variate analysis is inappropriate and meaningless. Furthermore, the authors were confronted by several additional difficulties which frustrated their desire to attribute performance on the ETS examination to several key factors. The group of 30 students who took the test had a mean grade point average of 3.23. Even in these days of grade inflation, the study group was clearly a very talented group academically. It seems plausible that the intrinsic academic excellence of this very special group of students might overwhelm, and thereby mask, the impact that any of the background variables would have on ETS exam performance. Before one dismisses this argument, it is appropriate to note that combined performance of these students placed them in the 98th percentile in terms of national performance based on mean scores. These are very bright and well-educated students. To dramatically illustrate the point, the authors' attempt to attribute performance on the ETS to key pedagogical exposures such as simulations, experiential exercises, etc. might be likened to

trying to attribute the performance of Harvard graduates, differentially and meaningfully, to whether or not they had had the benefit of pre-school.

With all of the qualifications that have been stipulated, in this day of "computer-generated" statistical analysis, it is tempting to "suggest" a few results. If they serve no other purpose, at least they indicate that the researchers had a research design that served as a blueprint for their analysis before they had to realistically confront the limitations of their data. Furthermore, the design may be useful for future research, particularly if one is interested in constructing meaningful output measures in accordance with the new AACSB accreditation guidelines. To that end, the student test scores on the ETS were dichotomized to form a new grouped variable based on each student's performance as compared to the median for the group of 30 students. Several bi-variate tables were constructed to see if significant relationships arose. For example, a table was generated which relates prior course exposure to simulations versus the grouped test score (GrScore). The results of this analysis appear as Table 1.

TABLE 1
RELATIONSHIP OF EXPOSURE TO SIMULATIONS AND GROUPED TEST SCORE

	Rows: SIM?	Columns: GrScore	
	Test score below median	Test score above median	All
No Simulation Experience	4	2	6
Prior Simulation Experience	11	13	24
All	15	15	30
Chi-Square = 0.833, DF = 1, P-Value = 0.361		2 cells with expected counts less than 5.0	

The obvious deficiency associated with limited degrees of freedom is apparent. Be that as it may, the P-Value of .361 suggests no evidence of a relationship.

Table 2 presents the results of a bi-variate table generated to investigate the potential impact of internships on the grouped test score (GrScore).

TABLE 2
RELATIONSHIP OF EXPOSURE TO INTERNSHIPS AND GROUPED TEST SCORE

	Rows: Intern?	Columns: GrScore	
	Test score below median	Test score above median	All
No Internship Experience	7	9	16
Internship Experience	8	6	14
All	15	15	30
Chi-Square = 0.536, DF = 1, P-Value = 0.464			

In this comparison the P-Value of .464 also suggests no evidence of a relationship.

In order to provide at least a crude check on the validity of the test scores, an analysis was performed to consider the relationship between the grouped test scores and the student grade point average (GPA). Again in this case, to facilitate analysis the GPA was dichotomized based on performance with respect to the median. The result of this process generated a variable called *GrGPA* (grouped GPA). Table 3 presents the results for the bi-variate table that relates grouped test score (GrScore) to grouped GPA (GrGPA).

TABLE 3
RELATIONSHIP OF GROUPED GPA AND GROUPED TEST SCORE

	Rows: GrGPA	Columns: GrScore	
	Test score below median	Test score above median	All
Below median GPA	11	5	16
Above median GPA	4	10	14
All	15	15	30
Chi-Square = 4.821, DF = 1, P-Value = 0.028			

The P-Value of 0.028 presents the reassuring evidence that there is something more than chance operating in this instance.

DISCUSSION

Obviously the paucity of data severely limits our ability to provide a meaningful analysis. Several years of similar size study groups, or somehow "bribing" --or even coercing-- a significantly greater proportion of our students to partake in this event, could improve upon the study's validity. This doesn't solve all of the difficulties though.

The students at the authors' institution do not comprise a reasonable representation of the overall college student body. They are too homogenous in terms of their ethnicity, race, family income and particularly their academic preparation. Average SAT scores and high school averages at the institution are generally quite high. As noted before, the group ranked in the 98th percentile on the ETS test, putting them on the extreme of the bell shaped curve.

Additionally, as noted in the responses to the questionnaire, all of the students had completed a course that involved a case study and over 80% of the students had taken a course in which either computer simulations or experiential exercises had been used. These results probably are representative of the authors' institution, where the pedagogical focus in many of the courses is on experiential learning techniques. Unfortunately, this presents the authors with another limitation in using *only their* institution for the study.

It appears that a meaningful test of the hypothesis requires a collaborative study with another institution(s). The need for a larger, more heterogeneous group of students is obvious. The study group should better represent the full spectrum of academic preparation – from the very bright to the not so very bright. A more demographically diverse student group would also enhance the validity of the study. Perhaps some of our ABSEL colleagues would welcome this opportunity for a multi-institution research project?

CONCLUSIONS

The study presented is an interesting blueprint for further research. The AACSB has determined that outcome assessment is an integral--albeit somewhat controversial--part of the continuous improvement process for business schools. The ETS examination provides one venue for measuring and assessing output. The opportunity to combine the administration of the exam with research on learning should be capitalized on. The compilation of a "home-made" test and a longitudinal study of this test compared to the ETS exam also provides a unique research situation. This may become a critical research issue if the AACSB determines that the ETS or some similar standard, national test is the ***only way*** to measure output!

The ability to expand the study group in terms of both size and diversity would certainly provide more meaningful results. Additionally, the authors recognize the need to look at the results on a subject by subject basis rather than using the aggregate mean for the test. It may be that there is no relationship between the exposure to experiential learning techniques and the mean score on the test because the exposure to these techniques is occurring predominately in one or two subject areas. An exploration of the possible relationship on a subject by subject basis is warranted.

REFERENCES

Available from authors upon request.