

Developments in Business Simulation & Experiential Exercises, Volume 15, 1988

EDUCATIONAL TESTING WITH THE MICROCOMPUTER

Dale L. Varble, Indiana State University
Bob E. Smiley, Indiana State University

ABSTRACT

The objective of this paper is to analyze software designed to aid in educational measurement. Viewed as central to education, measurement, whether carried out for evaluation or instructional purposes, rightfully requires considerable attention from the instructor. Student course evaluations depend on good measuring instruments. All too often instructors are criticized for their tests with comments like, "Real Picky" and "If you didn't happen to read one sentence you were penalized," in their student evaluations. The availability of micro computers has encouraged the development of software that both reduces the time necessary and also improves the process as normally conducted. There are, however, some limitations to software use and some measurement areas where software has not been fully developed. For example, the software may lock the instructor into a predefined instructional pattern not allowing for the instructors' own individual style.

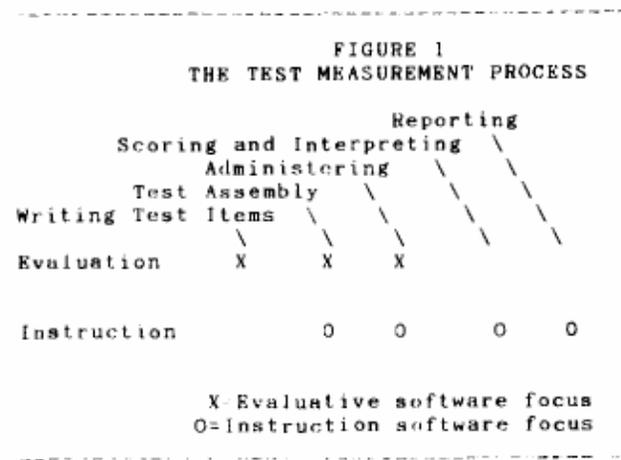
MEASUREMENT IN EDUCATION

Education involves ranging the behavior of learners in directions considered desirable: an increase in knowledge, skills, and problem solving abilities or a change in attitude. To determine how much change has occurred, measurement is necessary. The most common form of measurement in college business courses is a test. The test provides information. How does the learner perform after instruction? How does the student compare with other students? What instructional techniques worked best? Are examples of questions that we typically would like to answer from testing? The information obtained and its usefulness, however, will depend upon the instructor's objectives, the test items, test assembly, test administration, scoring, interpretation and reporting of results.

The process of testing is shown in Figure

The uses of testing can be thought of as stemming from two primary objectives: instructional or evaluative. Instructional uses of tests recognize the role of measurement in bringing about the desired change. Measurement may show the inadequacy of present behavior. Understanding the inadequacy of present behavior and the significance of how may be enhanced by measurement results. Measurement may indicate existing behavior patterns of individual learners. It may also provide the basis for the teaching and learning of specified behavior.

Objectives of Measurement



Evaluative tests attempt to determine the extent of behavior change. Measurement in an evaluative format should provide for and encourage self-evaluation, it may reveal major types of inadequate behavior and the contributory causes. Evaluative measurement may also permit and encourage individual initiative. A test may have both evaluative and instructional objectives. In college business classes, however, evaluation seems to be emphasized. Instruction occurs in a lecture format followed by a test. The test results are given to the students but rarely does the instructor have or take the time to go over the test in detail with the students. Neither is the student permitted to keep the test to analyze performance.

Evaluation and Software

The measurement software presently on the market has either an evaluation or instruction orientation, a limitation in this software. Evidently software that can be used for either evaluation or instruction has not yet been written.

To illustrate both positive and negative aspects of test software, several software packages are analyzed. MICRO EXAMINER 11, McGraw-Hill Publishing Company, is typical of the software from textbook publishers. EXAM and TEST GENERATOR are typical software from programming firms. MICRO EXAMINER is free to adopters of McGraw-Hill textbooks. EXAM is \$4995 and TEST GENERATOR is \$99.00. The three also differ in strengths and weaknesses with regard to the testing process as discussed below. All three, however, as with all evaluative test software, are strongest in the first three stages of the test process: test items, test assembly and administering.

TABLE I
EXAMPLES OF SOFTWARE AND THEIR PARAMETERS

	Test Bank Provided	Add Questions	Edit Questions	Delete Questions	Type of Questions	Question Selection	Random Questions	Randomize Foils	Randomize Within Question Type Only	Number of Test Versions Possible	Test Instruction Location	
Micro Examiner II	Y	Y	Y	Y	MT	I	Y	N	?	2	Top	
Microstudy	Y	N	N	N	M	R	Y	N	Y ¹	Many	-	
Test Writer	N	Y	Y	Y	MTEF	R-I	Y ¹	Y ¹	Y ¹	?	Top	
Tester	N	N	N	N	MTEF	R	Y	N	?	Many	-	
Archive2	N	Y	Y	Y	MTEF	R-I	Y	Y	Y	Many	Top ²	

Y = Yes
N = No

1. Multiple choice questions only
2. Test instructions at top and before each question type

R = Random question selection
I = Instructor questions selection
M = Multiple choice questions
T = True-False questions
E = Essay questions
F = Fill-in questions

A recent entry into the marketplace, .archive/2, is provided free of charge to adopters of Heath Co. texts. This exam package provides for all the question selection formats that anyone might want. However, it points to a basic problem with exam software. There appears to be little standardization in the industry. By this, we mean that the operator must review the literature on the package to determine not only what is available with that package, but also precisely how to operate that package. With the MICROTEST package, the operator only need to type Microtest to enter the program. With Arc-hive/2, they must enter Archive/h: to additionally name the deck for the question source location. While not a great change, to a department secretary who prepares many exams from many sources for many instructors this can be monumental form of wasted effort and time until the very subtle change is discovered. The Archive/2 program is also cumbersome because one must only select questions from one of three question file disks prior to moving on to the next. This forces all forms of questions from chapters 1 through 8 to be grouped together and then those from chapters 9 through 17 grouped and then those from chapters 18 through 24. This appears to be because the originator of the program reads the test bank into the memory and then selects and manipulates the questions. With the other programs, the selected

questions are real From disk, stored, and then manipulated after all the questions have been selected. This provides a level of ease of operation which is much better than the Archive/2. With arm agreed upon standard methodology, these problem areas would be removed and satisfaction with the test packages would improve greatly.

Test Items

The Test quest inns (an either be written or selected from an existing test bank. ARCHIVE/2 and MICRO EXAMINER? II are oriented toward selection or questions from a test bank, Publishers provide test banks with their software. The ease of item selection is the real strength of most publishers software. Indeed the ease of item selection may Improved teaching as increasing the number and frequency of tests. A benefit of more testing according to one research study, causes more anxiety but more effort to learn. With MICRO EXAMINER II, the instructor selects questions from the hard copy of the test bank and then enters a Four digit identification number for each question to include the question on the test . The test writer program randomly selects questions for the instructor to review and then allows the instructor to

Developments in Business Simulation & Experiential Exercises, Volume 15, 1988

de-select as well as select from these questions. MICRO EXAMINER II was designed to make entering, correcting and keeping track of the number of four digit item identification numbers easy. Because of its limited orientation, MICRO EXAMINER II is very easy to use if the instructor only wants to select existing questions from a test bank. The software does not give any assistance, however, in recording data on test items such as level of difficulty. All item record data has to be maintained in the hard copy of the test bank. Clearly it cannot be assumed that publisher provided multiple choice tests are properly designed. It would be especially useful if item evaluation was provided with the test bank. None were found, however, which could do this.

Writing and editing questions with MICRO EXAMINER II is more difficult than with EXAM or TEST GENERATOR. Neither EXAM nor TEST GENERATOR come with a test bank therefore their orientation is on building and maintaining a test bank. EXAM has word processing like functions that make composing questions on the computer possible. EXAM also permits an entry on question difficulty. Both EXAM and TEST GENERATOR randomize multiple choice question answers thereby overcoming the possible problem of position biases and patterns in correct response placement. Many instructors prefer not to have the correct response in the first position. They run out of good ideas, however, before the fourth and fifth alternatives. Thus the correct answer appears in the second or third position. Students who recognized the bias and patterns are at an advantage.

EXAM permits items to be of the multiple choice, true-false, matching, short answer, essay and fill in the blank format. Some software will handle multiple choice and true-false formats only. Some software will not accept questions or answers over a specified length. The first edition of TEST GENERATOR would not accept multiple choice answers longer than three lines. In strategy or policy courses where the questions and answers tend to be lengthy, size limitations are a prohibitive weakness.

Test Assembly

Physically putting the test together for distribution is test assembly. Print controls in the software make the difference in how a test is physically assembled. The headings, margin control, scrambling of questions, scrambling of multiple choice answers, paper size, numbering of questions and pages are some of the more important print controls. MICRO EXAMINER II has very limited print controls. The header at the beginning of the test can be controlled. Subheadings for each question type cannot be controlled by the instructor. MICRO EXAMINER II also permits two versions of the test, scrambling questions but not the answers. TEST GENERATOR allows for four versions of an exam but it scrambles both questions and multiple choice answers at the same time. It also scrambles across question type rather than within question type. Scrambling across questions may make scoring more difficult of each question

type has a different value. At the same time, scrambling across question type may make the test more interesting for the student. EXAM permits considerable control over assembly. Most software assembles one copy of each test version; multiple copies have to be generated using noncomputer duplicating techniques. MICROTST, however, will produce a copy for each student. For lengthy test printing a test for each student is an inefficient use of the low speed printer, but for quizzes the feature may be very useful. EXAM also prints answer sheets for the students.

Administering

The environment in which the test is administered/taken is influenced by the physical test. Cheating is more likely when a stressful, pressure-packed climate exists. With more than one version containing scrambled questions and answers, cheating by looking at a neighbor's paper has little possibility of succeeding. The psychological climate with scrambling improves. Students can relieve their test anxiety by taking momentary visual breaks by looking around without fear of appearing to cheat. Instructors can relax their vigil. MICRO EXAMINER II as previously mentioned scrambles only the questions while TEST GENERATOR and EXAM scramble both questions and answers. The instructor does have to hand out the test so students seated next to each other have different test versions. Students, when they realize that students sitting next to them have different tests, also realize quickly that to do well studying is required.

Scoring and Interpreting

The evaluative test software offers very little aid in scoring and interpreting an exam; all print an answer key, but nothing more. Scoring, interpreting and reporting are major weaknesses in the evaluative software. The next major developments should be in these areas. Progress, however, waits on the connection of an inexpensive and easily used data entry device to the microcomputer. Such a device would allow scoring of the exams and evaluation of the questions. The software could then be programmed to calculate the Kuder-Richardson indexes of reliability the variance, the mean, standard deviation, frequency distribution of scores and more.

These calculations are especially important from analyzing and comparing the different versions of a test. To be perceived as fair by the students, the test versions should have very similar statistics. The item analysis data such as difficulty, discrimination and answer distribution also helps improve the questions. By having difficulty and other item statistics stored in the test bank with the questions, test question selection will also improve.

Reporting

TEST GENERATOR has a "mapping" feature that aids in reporting scrambled questions results. A printout indicates how questions

Developments in Business Simulation & Experiential Exercises, Volume 15, 1988

in different versions are related. All students can focus on the same question at the same time when questions are discussed. But more is needed in reporting scores. Students want to know what their scores mean. The current software offers nothing in norm-referenced or content-referenced data interpretation. Being able to tell a student why he/she obtained the score improves the probability of learning.

Instructional Software

Instructional measurement, as stated earlier is intended to help in the learning process: to assist in bringing about the desired change in the student. Instructional test software may, therefore, be part of a more integrated system that encompasses the total teaching process. All instructional software to date are interactive and described as computer-assisted-instruction.

Several Programs are Described

MICROSTUDY, free to textbook adopters, programs the computer as a medium for presenting a study guide. MICROINSTRUCTOR, on the other hand, is an integrated \$995 program designed to assist in instruction--" tutorial, review, case study simulation, drill-and-practice exercises--all with full-color graphics. MICROINSTRUCTOR also is programmed to assist in construction and presentation of displays through computer-controlled videotape or videodisc. TESTER, at \$39.95, is an inter-active program designed to be used with an existing test bank.

Writing test items is allowed with some software. MICROINSTRUCTOR and ARCHIVE/2 permit the instructor to write any type test item. TESTER does not allow questions to be written but will accept questions written from a companion program TEST WRITER. MICROSTUDY accepts questions only from the study guide bank

Random selection of test questions is typical of the instructional software. TESTER, for example, permits either full tests or quizzes but randomly generates both from the test bank. Selection of questions by the instructors not possible.

Administration

At the software presents the test material on the monitor screen. Therefore, the student must have access to a computer and monitor. Paper presentation is not possible. In learning centers with large numbers of computers, access should not be a problem. For most instructors and departments, however, the availability of a computer for large numbers of students can definitely limit use.

Creating interesting presentation has been a weakness of the

typical instructional software. An attempt at generating interest has been to personalize the presentation of material. For instance, after initially requesting the student's name, a correctly answered question may trigger the flashing of "Good job, John" on the screen. The test questions, however, most often are presented one at a time in a rather mundane fashion with no bells, graphics or whistles to create the excitement sometimes necessary for motivating. There are, however, programs now being written which take advantage of the graphic and sound generating capabilities of today's micro computers. They usually are pictures of explosions and rockets shooting into the air. As programmers become satisfied with the basic evaluation process, they will add these graphic features and sound to make the level of motivation even higher.

Scoring and Interpreting

The instructional test software used with study guide questions provides only the correct answer to the student. Slightly more advanced software calculates the score for the student. Even more advanced software allows the questions to have different value weights, scores the test and conducts a scores analysis.

The more advanced software allows the instructor to generate a variety of reports, student performance, answer records and grades. The more integrated into the teaching process instructional software is the more likely reports will be a part of the program. All CAT software provides immediate feedback to the student, a very desirable feature in the Learning process. Quick feedback helps students evaluate their progress or lack of it while there is still time the academic term to remedy differences.

SUMMARY AND CONCLUSIONS

Instructional and evaluative measurement has not been combined in software programs. Software tends to focus on either instruction or evaluation. If focused on instruction, the elementary software uses the computer merely as a medium for presenting questions. A study guide booklet would be more effective and efficient. The instructional software that incorporates the whole testing process into the program is completely computer assisted instruction. The computer has to be used, to separate and return to the computer during the testing process is not possible with available software.

The evaluative measurement software focuses on the first three steps of the testing process: item selection or writing, test assembly, and administration. Scoring interpretation and reporting receive minimum consideration in the programs. Most software does not allow for feedback from scoring, interpreting and reporting from other sources. Overall test software has significantly aided in the testing process. Too often, however, the software focuses on one area of the test-

Developments in Business Simulation & Experiential Exercises, Volume 15, 1988

ing process and locks the instructor into a predefined testing procedure in the other areas of testing.

When truly developed, the software will permit the evaluation for the student, provide feedback and allow versatility in generating the test instrument in either questions, answer positions, or any combination of them.

Further, this software will permit the micro computer to read and interpret the student responses and have an evaluation of the question available in order that the instructor and the student might better organize the student evaluation instrument for the maximum benefit in the learning process.

REFERENCES

1. Archive/2 (Lexington, Massachusetts: D.C. Heath and Company, 1987).
2. Bozinoff, Lorne, "An Evaluation Methodology For Marketing Tests." 1983 AMA Educators' Proceedings, Patrick E. Murphy, et. al. (eds.) (Chicago: American Marketing Association) pp. 155- 158.
3. Brown, Frederick G., Measuring Classroom Achievement. (New York: Holt, Rinehart and Winston, 1981.)
4. op. cit. Brown.
5. Bwyer, Robert F., "Teaching with Tests: Polemics, Propositions, Payoffs' in Neil Beckwith et . at. (eds) 1979 Educators' Conference Proceedings (Chicago; American Marketing Association) pp. 116-119.
6. Jackson, Donald W. Jr., Nabil Razzouk and Louis H. Grossman. "Evaluation in Marketing Education: The Case of the Final Exam." Developments in Marketing Science, Vifnay Kothari (ed). Proceedings Academy of Marketing Science, 19R2, pp. 257-260.
7. Madden, Charles S., and Denise T. Smart. "Alternatives to Traditional Faculty Evaluation Measures: A Comparative Analysis." AMA Educators' Proceedings, Patrick E. Murphy e.a. (eds), (Chicago: American Marketing Association) pp. 164-168.
8. Micro Examiner 11. (New York: McGraw- Hill Publishing Co.), 1985, and Exam, (Denver, Colorado: Brownstone Research Group, Inc.), 1985.
9. Microinstructor. (St. Louis, Missouri: Moshy Times Mirror, 1985).
10. Microtest: A computerized test preparation system (1984), Omaha, Nebraska: Delta Software, Inc.
11. Payne, David A., The Identification and Measurement of Learning Outcomes. Waltham, Massachusetts: Blaisdell Publishing Company, 1968.
12. Popham, W. James, Modern Educational Measurement. (Englewood Cliffs, New Jersey: Prentice-Hall, Inc.) 1981.
13. Ibid.
14. Test Generator(1985), Terre Haute, Indiana: Sycamore Press, Inc.
15. S.O.V.E. Tester (1984), Ashland, Oregon: Southern Oregon Video Enterprises.
16. S.O.V.E. Test Writer (1985), Ashland, Oregon: Southern Oregon Video Enterprises.