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’ arid “If you didn’t happen to read one sentence you were penalized,” in their student evaluations. The availability or 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.

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 lo 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.
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 Archive/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 in terms 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 files after selecting all the formats of questions from chapters 8 to 17 to be used, then select questions from chapters 18 through 24. This forces all selected questions to be real-time from stored disk, manipulated, and then reviewed. This is much better than the MICROTEST or ARCHIVE/2 with its random selection of questions from a test bank for the instructor to review. The test writer program randomly selects questions for the instructor to review and then allows the instructor to

<table>
<thead>
<tr>
<th>Test Instruction Location</th>
<th>Number of Test Versions Possible</th>
<th>Randomize Within Question Type Only</th>
<th>Randomize Foils</th>
<th>Question Selection</th>
<th>Type of Questions</th>
<th>Delete Questions</th>
<th>Add Questions</th>
<th>Test Bank Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro Examiner II</td>
<td>Y Y Y Y MT I Y N ? 2 Top</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microstudy</td>
<td>Y N N N M R Y N ? + Many</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Writer</td>
<td>N Y Y Y MTEF R-I Y Y ? ? Top</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tester</td>
<td>N N N N MTEF R Y N ? Many</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archive/2</td>
<td>N Y Y Y MTEF R-I Y Y ? Many Top</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 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 a hard copy of the text book 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

Dealing with multiple choice questions.

Some software will not accept questions or answers over a specified length. The first edition of TEST GENERATOR formats only. Some software will handle multiple choice and true-false matching, short answer, essay and fill in the blank format. EXAM permits items to be of the multiple choice, true-false, matching, short answer, essay and fill in the blank format. 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 Kudem-Richardson indexes of reliability the variance, the mean, standard deviation, frequency distribution of scores and more.

These calculations are especially important for 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
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 interactive 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 rest 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 is not possible.

Administration

AL! 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 students 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 to remedy the academic term.

SUMMARY AND CONCLUSIONS

Instructional and evaluative measurement has not been combined in software programs. Software lends to focus on either instruction or evaluation. If focused on instruction, the elementary software users 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, he software focuses on one area of the test-
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


13. Ibid.

