R E P O R T R E S U M E S

ED 014 288 JC 660 494

COMPUTER TECHNIQUES FOR WEEKLY MULTIPLE-CHOICE TESTING.

BY- BROYLES, DAVID
RIVERSIDE CITY COLL, CALIF.

PUB DATE 66

EDRS PRICE MF-$0.25 HC-$0.36 7P.

DESCRIPTORS- *JUNIOR COLLEGES, *TESTING, *COMPUTER ORIENTED PROGRAMS, MULTIPLE CHOICE TESTS, POLITICAL SCIENCE, ELECTRONIC DATA PROCESSING, STUDENT TESTING, *TEST CONSTRUCTION, TEST VALIDITY, ITEM ANALYSIS,

TO ENCOURAGE POLITICAL SCIENCE STUDENTS TO READ PROPERLY AND CONTINUOUSLY, THE AUTHOR GIVES FREQUENT SHORT QUIZZES BASED ON THE ASSIGNED READINGS. FOR EASE IN ADMINISTRATION AND SCORING, HE USES MARK-SENSE CARDS, ON WHICH THE STUDENT MARKS DESIGNATED AREAS TO INDICATE HIS NUMBER AND HIS CHOICE OF ANSWERS. TO EMPHASIZE THE VALUE OF CONTINUED HIGH LEVEL PERFORMANCE AND TO INFORM STUDENTS OF THEIR CURRENT STATUS, CUMULATIVE SCORES ARE REPORTED. ADVANTAGES OF THE PROCESS INCLUDE (1) AUTOMATIC SCORING AND REPORTING WITH LITTLE EFFORT OF THE INSTRUCTOR, OTHER THAN PREPARATION OF A CORRECT MASTER CARD; (2) USE OF COMPUTER DATA STORAGE TO COMPILE CUMULATIVE SCORES; (3) AUTOMATED COMPUTATION OF GRADES BASED ON THESE AND OTHER EVALUATIVE SCORES, AND (4) ITEM ANALYSIS FOR IMPROVEMENT OF THE TESTS. BY RELIEVING THE TEACHER OF MUCH OF THE MECHANICAL ROUTINE ASSOCIATED WITH TEST SCORING, RECORDING, AND ANALYSIS, THIS SYSTEM PERMITS MORE FREQUENT AND EFFECTIVE EVALUATION. (WO)

COMPUTER TECHNIQUES FOR WEEKLY MULTIPLE-CHOICE TESTING

DAVID BROYLES

RIVERSIDE CITY COLLEGE

Spring, 1966

This paper reports a process of computer-processing tests which the writer has used to some advantage in an instructional situation where weekly multiple-choice testing was advisable. The process may be useful to others either to meet a similar instructional need or for other reasons. The report is brief, but further details can be furnished on request. Details concerning machine processing and computer programming have been held to a minimum, since they would not be useful to the typical reader.

The instructional problem which caused the author to pursue a new testing process was the following: Students in Political Science (American Government) are disposed to take the course lightly since it is not usually associated with a "major" program and for other reasons. This means that they do not prepare adequately for class on a day-to-day basis. As a result, they miss what is perhaps the most valuable part of the course - the continuing training in investigation of, and discourse about political problems. This is training which is most vital to the pursuit of intelligent adult citizenship. For training of this type to be effective, the student should feel the urgency of preparing much as he would for a language class. Except for the necessity to impart this training, there

would be no objection to any study method which the student wanted to follow. That is to say- if the objective were simply to impart information about the mechanics of government, it would not really matter whether this information was acquired all in one night before the final examination or even before the course had started.

In accordance with this training-in-discourse concept of the function of the course, the midterm and final examinations were made essays in which the student was challenged to convince someone else of his view on an important political matter. To prepare for this, the students were rewarded well for participation in class discussion and debate. They were not rewarded merely for class attendance or for participation, however, since the former was consistent with passivity and the latter was very difficult to evaluate. Instead, they were rewarded for preparing for class discussion and debate on the assumption that actual participation would follow. Students were assigned materials which were not descriptive but frankly argumentative. When readings of this character are assigned, it is not too difficult to test proper comprehension of them, as is illustrated by the following example:

> The first chapter of your reading XYZ takes the position that all redheaded men are warlike. This position is supported by:
>
> 1. Statements showing that no blackheaded men are warlike.
> 2. Examples of redheaded warlike men
> 3. Arguments from genetics
> 4. None of the above, since the position is not taken

Another example would be:

> Of the men you have read this week, which would be likely to favor Medicare for the reason that it would correct a social injustice caused by the industrial revolution:
>
> 1. Hamilton
> 2. Madison
> 3. Jay
> 4. Paine

Such questions, if given frequently on assigned readings, encourage students to read properly and continuously. This in itself is valuable training. It also results in improved classroom participation, since students are better prepared, although the desirable 100% participation level is certainly not achieved by this method. Participation based upon this kind of preparation is of greater advantage to the student than mere recitation, since there is greater incidence of the position-taking and position defending which is a necessary characteristic of true discussion and argumentation.

It was thought desirable to emphasize the value of continuous high level performance and to keep reminding students of their progress on a cumulative basis. Therefore, when the second test was given it was to be reported back with the first test and in fact scored together with the first test. Consequently students were graded on the basis of percent right of the total taken to date. If, for example, on a first five-point quiz, a student got four right, he received a score of 80. If, on the second five-point test, he got only three right, he received a score of 70 (7 out of 10 not 60 (3 out of 5)). And so on.

It was also considered advisable to report back correct answers to each student along with his own answers. In this way, the student could check his work against the test which he had been allowed to keep, and could learn from his own mistakes.

IBM cards are used to collect answers to objective questions. Blank cards are distributed to students along with electrographic pencils. Each card contains spaces in which to mark a student number and other spaces in which to mark answers to almost 100 questions. To indicate his student number, the student draws lines across five "bubbles" which surround digits of the student number. To answer a question, he draws a line across a "bubble" which contains the number of the answer which he has selected. One advantage of the use of cards is that all processing is automatic. From the "bubble" markings alone, data processing equipment can identify the student, grade his paper and issue a report by name or by student number on cards (one per student) or on a listing. The grading is accomplished by comparing student answers with those indicated by the instructor on a specially designated master card. The data processing equipment is able to "remember" the students previous performance and to report his present score on the above-outlined cumulative basis. This ability of data processing equipment to "remember" previous performance also makes it possible, if the instructor so desires, to keep a performance record for each student (or duplicate performance records) in a centrally located data processing file. This file would be available in case the

instructor's records were lost. It could also be used to automatically compute a final course grade on the basis of weighting or other instructions from the instructor. In other words, the final course grade could be assigned by data processing with the final examination forming one element and prior performance providing other elements.

One advantage, then, which a card answer form has over a paper answer form is that it can be processed and results can be accumulated with previous results all automatically with the instructor doing no more than making out the tests and marking a correct-answer card. But this is not the only, or even the major advantage. The card is also much more easily manipulated for purposes of test analysis. One analysis will give an indication of the efficiency of a question in selecting between good and poor students. Using the overall test scores, the machine ranks students and places them in quartiles or deciles. A effective question will be one where higher deciles or quartiles show more correct answers than do lower ones. A convenient index can be calculated for rating questions on this basis. One such index is calculated by taking the proportion of right answers to total answers achieved by the lowest quartile students and subtracting this from the proportion of right answers to total answers achieved by the highest quartile students. Thus, if 5 out of 10 students got the right answer in the highest quartile and 3 out of 10 in the lowest, then the calculation would be $\frac{5}{10} - \frac{3}{10} = \frac{2}{10}$ The result is always a number between plus

1 and minus 1 with negative numbers indicating that the question tends to benefit poor students more than good students. A supplement to this index can be usefully calculated. It shows the success of the question in terms of how difficult it was. If too many students missed the question (or if too many got it right) it does not do the job of discriminating between good and bad students which it should do. Selectivity and difficulty indexes when used together provide the instructor with a good indication of the usefulness of each individual item on his test. This evaluation can lead to changing subsequent tests for the better or it can be used to adjust the grading of the test for which the indexes are being calculated. The instructor can specify that certain types of questions (such as those where less than thirty percent or more than ninety percent of students get the right answers) be eliminated automatically before final grades are assigned for the test.

Another useful set of analyses can be obtained based on content analysis of each multiple-choice question. Frequently, multiple-choice questions are formulated on the basis of an assumption that one answer is best, another is almost as good, a third is likely to look attractive to a person who has no familiarity with the material at all, and a fourth is only for the random guessers. Such questions are good, since they engage the instructor in an analysis of how the material which he is teaching is affecting the students, and why students are likely to make errors on questions. Whether or not a question was

received in the manner intended is evident from a frequency count showing how many students got each answer and how many students in each quartile or decile got each answer. Such an analysis may show, for example, that the very best students in the class are answering the second best answer. The reasons for such a finding would be well worth investigating. Also, a single index (a coefficient of correlation) can be calculated to measure the instructor's success in predicting which answer is the best, second best, third best, and fourth best.

These are some analyses of test answers which the author has used, and which were made possible by using card answer sheets rather than the more usual paper answer forms. Other analyses are also possible. For example, a computer can scan stacks of student papers for patterns of wrong answers in such a way as to check for student cheating. The pattern of answer analysis groups students who have all or most answers the same, thus permitting the instructor to intelligently pursue the question of possible cheating. The results of these analyses would assist the teacher materially in administering frequent and helpful tests while unburdening him from much of the mechanical routine associated with test analysis, grading and recording grades.