DOCUMENT RESUME

ED 077 981 TM 002 791

TITLE [The PEDAGE System.]
INSTITUTION Toronto Univ. (Ontario). Dept. of Geology.
PUB DATE 22 Dec 65
NOTE 7p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Computer Programs; *Scoring; *Student Testing; *Test Construction; *Test Results
IDENTIFIERS *PEDAGE System

ABSTRACT

The PEDAGE system, at the present time, is a set of programs written in FORTRAN IV language for an IBM 7(computer. They are in the form of separate main programs with si ilar data formats, but will be modified in the future to the form of subroutines so that one or more can be called by main programs designed by the individual users. In this final form, the system will be available in FORTRAN decks, binary compiled decks, and binary compiled magnetic tape. At present, the separate programs are available for distribution in the first two forms. The programs already developed are for three functions: (1) to select and print a set of declaratory statements that are either true or false, (2) to score and analyze student performance in true/false tests, and (3) to evaluate the efficiency of the individual statements in scoring the students. (For related documents, see TM 002 778, 789, 790, 792-793.) (DB)

DEPARTMENT OF GEOLOGY
UNIVERSITY OF TORONTO
TORONTO, ONTARIO
CANADA

General information on the PEDAGE system of computer programs for scoring and analyzing methods of teaching and examining knowledge of factual material.

PEDAGE

December 22, 1965

THE PEDAGE SYSTEM

Introduction. Computer programs to score and analyze true/false tests were first tried in this department in a course of elementary geology in the 1964-65 session, supervised by F.G. Smith. The results were favorable and a more intensive trial was carried out in the first term of the 1965-66 session. The results were good and indicated that it may be feasible to use computers to do many of the operations involved in teaching. Programs and subprograms collected and maintained for this purpose are being classed as a system with the code name PEDAGE.

The PEDAGE system at the present time is a set of programs written in FORTRAN IV language for an IBM 7094 computer. They are in the form of separate main programs with similar data formats, but will be modified in the future to the form of subroutines so that one or more can be called by main programs designed by the individual users. In this final form the system will be available in FORTRAN decks, binary compiled decks, and binary compiled magnetic tape. At present the separate programs are available for distribution in the first two forms.

The programs already developed are for three functions, 1) to select and print a set of declaratory statements that are either true or false, and 2) to score and analyse student performance in true/false tests and 3) to evaluate the efficiency of the individual statements in scoring the students.

Purposes and general outline. The primary purpose in utilizing a computer to help in the housekeeping part of teaching of science and engineering is relieving the instructor of time-consuming drudgery such as marking examination papers. This in itself would justify using an available computer in testing large classes, but secondary benefits not so obvious as this one become apparent even in reconnaissance studies. The great speed of the machine makes it practical to develop hybrid teaching methods that approach the teaching machine ideal of a large but finite number of increments of learning, each increment being so small that the probability of mistake at each step is below some arbitrarily small constant.

We assume without argument that the traditional method of lecturing for one or two terms, followed by essay-type examinations, is not efficient in terms of concepts learned by the student or in terms of adequacy of sampling them with a small set of questions. We accept, also without argument, generalized results of trials of teaching machines that indicate their high efficiency. However, such machines are not generally available in sufficient numbers, and it may be several years before tested programs for a large number of courses become available. Therefore a practical procedure to increase teaching efficiency is to test concepts presented in lectures and text-books at short intervals throughout the term or terms of instruction and to dispense with final examinations. The optimum interval between tests is subject to investigation, but results in this department indicate that a test at the end of each instruction period is practical and efficient. For courses not based on a text-book, one test every two periods may be better.

We have concentrated our attention on the true-or-false type of examination of descriptive material in elementary general geology courses. This type of examination appears to be adequate for the purpose, but it puts a heavy load on the supervisor who must design for each test a set of statements that are in the correct range of difficulty and also unambiguous to the majority of the students. As a partial solution of this difficulty, some of the programs in this system allow storage of true-or-false statements about all concepts, terms, definitions, facts, etc., discussed in the course, and from this set a sub-set is selected by the computer and printed in formal manner for duplication for any one test. The supervisor specifies the subjects permitted, the range of difficulty, and the number to be selected. The computer picks out random sub-sets of the permitted statements, the first being used for the test, and others for supplemental tests, tests of absentees, and so on. Using this scheme, the number of possible statements can be increased year by year and improved by deleting, changing and adding statements to the set.

Assuming that a test is included in each lecture period of 50 minutes, the test itself should not take a disproportionate amount of the total time, but the best ratio of time in presenting and discussing concepts, to the time testing absorption of the same, could be worked out by the instructor. If many terms and relationships must be discussed, as in elementary courses, the test method cannot include time to write answers. From our experience, a test of 81 true-or-false statements using mark-for-true and blank-for-false coding, can be carried out in about 17 minutes. If the coding sheets or cards are premarked with the names of the students, these can be picked up as they enter the room. In addition to saving time, this decreases the chance of error in matching names and responses. A practical procedure is using prepunched mark-sense data cards for the coding, eliminating all human processing of the responses. This scheme is discussed more fully below.

An efficient examination procedure should include a rapid feed-back of the results to the students. Ideally, they should have the results before they leave at the end of the period. This is practical if quick access to a computer is possible, because the execution time of scoring programs is about two seconds, but most university departments do not have such a facility. However, a reasonable compromise is to have the results on hand before the next lecture period and the first part of the lecture could deal with patterns of poor responses. Some of the programs in the PEDAGE system print out reports to the individual students about their own patterns of performance, and others have output in the form of a report to the supervisor about class patterns of performance. This aspect of the system is under development at the present time.

Selecting appropriate statements for true/false tests. A tentative and fairly general computer program for random selection of a set of true/false statements is being developed and tested. The current best version is SLEXTF-M2. This selects a set of statements, permitted by a string of categorical descriptors and a range of difficulty, from a long list read from data cards. From this set are selected randomly one or more subsets of desired size, and the corresponding statements are printed in a formal way for the examinations. One page of the output contains the corresponding logical values of the statements for the supervisor.

A practical procedure is to obtain several equivalent but different examination sets at any one execution, and to use one for the main test and others for supplemental tests of absentees,etc..

The current list of statements for a first course in physical geology is not more than 500 and storage on punched cards is still feasible, but we will be storing statements (coding for search by programs similar to the above) on magnetic tape. This will require programs to handle loading, reordering, deleting, modifying, substituting, etc.. Development of these programs will be carried out early in 1966.

Scoring true/false tests and analyzing student response. The first programs were developed for the purpose of scoring true/false tests and analyzing in a simple manner the pattern of response. Program MARKTF-M3 is typical of this kind and it was found to be quite useful. The output consists of individual reports from the computer to the students and some primary statistics for the instructor. The declaratory statements used in the test are considered to contain four arbitrary categories of subject matter, mixed in any way, and part of the data is a matrix of the instructor's estimate of the loadings on each statement. The program makes subsidiary scores of the responses in the four categories and instead of giving the numerical results, prints out one of ten levels of advice about the results, for each of the four categories. The categories may be simply subject matter from chapters 1, 2, 3 , and 4 for example, but may be levels of logic such as 1) concepts and generalizations, 2) scientific hypotheses, 3) application to real things, 4) qualitative and quantitative facts, etc.. The advice is in the data deck and is selected by the instructor to apply to the class tested. A certain amount of facetiousness, sarcasm , irony, or even caustic comment can be employed without endangering the student-teacher relationship, because it seems that the inanimate computer decides what to print.

Scoring true/false tests and analyzing the test statements. Another set of programs was designed to make statistical analyses of the efficiency of each of the test statements in examining the performance of the students. Programs MARKTK-M5 and -M6 are typical of this set. The output gives numerical data which indicate whether the individual statements were good or bad for scoring and discriminating between scholars and drifters, also whether the statements were too easy, too difficult, or the wording was tricky in the sense of suggesting true if false or false if true.

The above two programs include an option which allows the user to accept regression slope parameters of each statement as weights on the responses. With this option, the performance of the students controls the scoring. One of these programs (M5) uses the slope parameters as generated, and some may be negative for very poorly worded statements. This causes quite a distinct partitioning of the student scores into two groups, each of which has a nearly normal distribution. The other program (M6) throttles the slope parameters between zero and unity. In effect, this puts some weight on the supervisor's knowledge of truth or falsity of the statements. The partitioning of the scores into two groups consequently is less in magnitude but is still distinct.

Data acquisition. Mark-sense data cards were found to be suitable for the true/false tests. A new form of these was devised to hold 81 bits of information on one side (IBM electrotype number 78326) and these are illustrated in the descriptions of individual programs. Usually, they are prepunched with the name and other data of the student so that there is no chance of mismatching test results and names. After any test, the instructor punches the prior current percentage standing into the card and any increment or decrement that is to be combined with the test score before combining with the prior standing to give a new current standing. The deck of student cards, and the instructor's control card of the same type, is put through a mark-sense punching machine and a copy of this deck becomes part of the data deck for the programs.

Projected developments. Computer programs for other kinds of analysis of true/false examinations have been written and others are planned. We will be converting these into a package of compatible subroutines and an option-specifying main program.

We plan to install a remote terminal of the IBM 7094 computer in a central location in our department. This will give us the facility of scoring and analyzing tests of knowledge within about three minutes and thus make it feasible to select a test set of statements, carry out the test, score the results, and analyze them within one lecture period and have about half the period left for discussion, presentation of new material, video recordings, etc..

We also plan to obtain one or more teaching machines for controlled tests of their efficiency relative to other teaching methods. Probably computer programs will be required to optimize the sequence, incrementation, recycling and so on of the teaching machine programs.