tanulmányok

194/1986

MTA Számitástechnikai és Automatizálási Kutató Intézet Budapest

Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete
Computer and Automation Institute, Hungarian Academy of Sciences
Исследовательский институт вычислительной техники и автоматизации
Венгерской Академии наук

СТАТЬИ КОНФЕРЕНЦИИ КНВВТ ПО АВТОМАТИЗАЦИИ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ НА ПЕРСОНАЛЬНЫХ ЭВМ

Будапешт, 5-9 мая 1986 г.

Том II.

KNVVT CONFERENCE ON AUTOMATION OF INFORMATION PROCESSING ON PERSONAL COMPUTERS

Budapest, May 5-9, 1986

Vol II.

Будапешт, 1986

Budapest, 1986

Tanulmányok 194/1986
Studies 194/1986

A kiadásért felelős:

DR. KEVICZKY LÁSZLÓ

Редактор:

ИШТВАН РАТКО

Editor:

ISTVÁN RATKÓ

ISBN 963 311 223 0

ISSN 0324-2951

87/65. Alfaprint

F.v.: Barabás Gábor

CONTENTS

Page

PERSONAL COMPUTER-AIDED TESTING AND TRAINING SYSTEMS

Radoslav Pavlov, Ruslan Mitkov, Avram Eskenasi

Insitute of mathematics, Sofia

ABSTRACT

In the present paper are examined the basic principles of achievement testing, the historical development of its computerization and mostly some computer-aided testing and training systems developed at the Institute of mathematics of the Bulgarian Academy of Sciences. The systems were developed on the personal computer "Apple-II". Certain advantages of the computer-aided testing and training (the latter realized efficiently only by means of computer), as compared to the traditional testing and training are examined, as well as future problems and tendencies of computer-aided testing and training.

ACHIEVEMENT TESTING

An achievement test is a systematic procedure for determening the ammont a student has learned. Testing has been around for centuries, for longer than most people today realize (ancient China). An widespread and rapid development of testing procedures and generally of theory of testing was noticed at the end of the 19th and the beginning of 20th century and the names of Rice, Thorndike and Binet deserve to be mentioned.

Achievment tests are widely used in education as a suitable instrument to measure someone's knowledge. The tests can be classified as essay, problem and objective tests. The objective tests can be scored more rapidly and reliably than either of the other types (especially preferable when large amounts of students are to be tested). Usually a test is a collection of test items. Objective items may be : true-false, multiple-choice, matching and short-answer.
It is certain, that multiple-choice test items are currently the most

highly regarded and widely used form of objective test items. They are adaptable to the measurement of most important educational outcomes - knowledge, understanding, and judgement ; ability to solve problems, to recommend appropriate action, to make predictions. It was the multiple-choice objective test items , that drew the attention of the informaticians in the development of the first computer-aided testing programs and systems. Where ever testing is concerned, we may speak also in terms of "training" (drill and practice), provided there is appropriate feedback.

LOOKING BACK : THE FIRST STEPS IN COMPUTERIZATION OF TESTING

The first computer programs(and later systems) were CTSS (USA, 1969) , MEDSIRCH (CANADA, 1970) , MENTREX (USA, 1973). These programs permitted item entry and storage (in the socalled "Itembank"), item bank maintenance, item generation and item scoring. In the following years in several countries (USSR, Japan, Bulgaria) were developed systems with more flexible and extensive possibilities. The systems TEST-1, TEST-2 and TEST-3 were developed in 1977 at the Institute of mathematics of the Bulgarian Academy of Sciences. At the beginning, the test items were kept on punch-cards and magnetic tapes. Later on, with the development of computer technology, it was possible to ensure direct access to the stored on hard discs test items. The appearance of personal computers with their new interactive (dialogue) and graphics possibilities gave a strong impact to the development of computer-aided education . In the field of computer-aided testing and training it was possible to think of improvements of all the stages connected with testing - test development, test administration, test scoring and reporting practices, item analysis and with training - the emergence of new training individual strategies, including some adaptive features.

These ideas were incorporated in the from us developed microcomputer testing and training systems.

UTEST AND UTRAIN - COMPUTER-AIDED TESTING AND TRAINING SYSTEMS

The systems can be described in two different ways : from viewpoint of the user-to be examined or trained (here called "student") and from the viewpoint of the user-examinator or trainer (here called "teacher"). From viewpoint of the student UTEST may be described as follows : the student sits in front of the microcomputer, receives brief instructions (several sentences on the screen) and is offerred a test, consisting of multiple-choice test items. The test item has a problem situation and several (3,4 or 5) alternatives, one of which correct and the others - distractors. The test item may also contain graphical illustration to the problem situation. The student is supposed to select the correct answer (choosing for instance "1", "2" or "3") or delay his answer. If after the generation of the test and its display on the screen the time is not up, the delayed items are offerred sequentially again. When no more items available, the results are stored in a file, further accessible to the teacher.

UTRAIN is a modification of UTEST on the basis of drill and practice mode. The student may have immediate feedback after answering each item ; the time, determined for examination may be adjusted according to the individual pace of work ; there is no assigning notes or mastery or nonmastery to each student, but just drill and practice covering desired subject areas.

From viewpoint of the teacher UTEST may be regarded functionally: UTEST consists of four larger subsystems : text base maintenance (text base editor), graphics base maintenance (graphics editor), test generation and examination and results processing. These systems are functionally connected through different files.

UTEST uses floppy discs and this brings certain restrictions of the data base size and technological complications of the results processing (additional complication of the latter is that the computers are not connected in net).

Special database maintenance means are elaborated for the examinator. Text and graphics are stored separately and supported in different ways. The text base maintenance means (text editor) represents an elementary menu and no preliminary knowledge is necessary to correct, update or eliminate text. The graphics base maintenance means (graphics editor) enables easy creation of various colour figures, which occupy minimal memory space and their maximally fast execution. To make use of the graphics editor is not hard to learn. The teacher is in position to require the generation of a test from the data base with determined characteristic features - number of test items, subject areas, global difficulty, time for test solution. Each test item is characterized by "difficulty weight" and belongs to certain subject area.

It should be mentioned, that UTEST is one of the not so many really functioning testing systems. We have managed to realize UTEST in several different subject areas. The most popular concrete version of UTEST is KATEST, a testing system for administration of the motor vehicle driver's exam (4) . KATEST is a menu-driven testing system. Its main menu looks like that :

1. Examination
2. Results processing
3. Text base maintenance
4. Graphics base maintenance
5. Copying of diskettes
6. Exit

Option 5. is necessary because of the special protection measures taken (KATEST works under completely different, specially elaborated DOS). They were a number of experiments, carried out with KATEST. We were glad to establish , that KATEST was friendly accepted by most of the examinees.

A more detailed functional scheme of KATEST is given in the appendix (page 13).

KATEST was not the only concrete realization of UTEST. For the students of Sofia University we developed testing systems for biology, physics, mathematics and law. We are working now on some perfection of UTEST, which is going to bue used by the Sofia Council of People's Education. We plan to include some powerful information subsystem , giving any information concerning the "achievement history" of each student. Furthermore, item analysis should be carried out by the system.

The authors are working now on a project, which will connect a group of computers in net and use common big capacity disc. The present paper will not discuss the advantages of such solution.

Comparing UTEST with traditional testing, we may state the following advantages in favour of UTEST :

- The number of tests offerred on traditional answer-sheets is not very large. This enables sometimes students learning tests by heart in advance and so arises the problem of test security ;in UTEST the test items are selected within each subject area by random generator - it means, that provided the data base is large enough, practically an unlimited number of distinct tests are produced, not known in advance by students.

- It is very difficult and expensive to change the tests in traditional way. Only a single change of an item, resulting from certain pedagogical or administrative reasons leads to the print of numerous tests. In UTEST all is simply done : the respective item is retrieved and by means of the text editor corrected.

- a similar problem arises when the structure of the test should be changed.In UTEST the structure of each test may be predetermined by the teacher.

- the advantage of UTEST is similarly seen when the global weight of the test should be changed

- the subjective factor is in UTEST fully eliminated.

It should be also mentioned, that UTEST is independent on the subject area, i.e. universal with regard to it. The user-friendly menu enable teachers even with little or no knowledge in programming to create their own data base.

In UTRAIN there exists the same functional structure. However UTRAIN has more options connected with the individual preparation of each student. The student is allowed to select a desirable teststructure for the drill and practice, to select the time for it, to select the mode of feedback. These options can be also selected by the teacher. Furthermore UTRAIN may analyze the level of knowledge of each student, instead of simply assigning a corresponding note to it.

ADAPTIVE FEATURES

The authors are trying to incorporate some "adaptive features" in the systems. Adaptivity is realized in examination differently from drill and practice. Essential in both cases is the subject determination of the test items. After some analyses and hesitations the authors preferred the hierarchical classification scheme (HKS) as more suitable than describing each test items by means of key words. The HKS contains 6 levels, and according to the properties of this type of languages each subject area is described by code (from 1 to 6 digits) and each subject area coded by k digits represents a subarea of each area, coded by the leftmost m digits of its code ($m < k$). Each test item belongs exactly to one subject area. Here is a simple example of it :

1. Mathematics

1.1 Elementary mathematics

1.1.1. Arithmetics

....

1.2 Higher mathematics

1.2.1.........

1.2.10 Probability theory

1.2.10.1. Random events

1.2.10.2 Random variables

....

2. Informatics (computer science)

2.1. Programming

2.1.1 Programming languages

....

3. Mechanics

The authors are acquainted with some attempts aimimng to set up adaptive CAI systems. Some of these systems are based on creating a model of the student and adapting the CAI (computer-aided instruction) in connection with this model . Others make use of sophisticated statistical theories for estimating student's ability. In this paper the authors describe their aim to obtain adaptivity in two strictly defined and not so complicated sides of the education - examination (testing) as well as drill and practice (training). What is more, they aim at obtaining this adaptivity by maximally elementary and user-accessible means.

The simplest adaptivity means in UTEST represents the possibility to postpone the answer of any test item. We can talk in this case about adaptivity and psychological adjustment of the student. The experiments, carried out, show that this possibility is frequently made use of. Unfortunately the authors do not claim any final conclusions, for too large statistical information is not available. This remark is also valid for the rest of the examined problems here.

Another adaptivity strategy is tealized in UTRAIN based on HKS. The teacher determines a structure of the drill and practice (just like in case of examination) : subject areas A_i, n_i , such that $n_i \leq card(A_i)$, total difficulty weight of the items of each area - M_i. Besides, a treshold number T_i for each area ($T_i \leq M_i$) and a number of loops N_i for each area are added. In case the student reaches at least T_i of given area, he receives items from the next area A_{i+1} . If the treshold number is not reached, the test items are presented once again on the screen with indication of the correct answers. After reading the correct answers, another n_i Items of A_i are generated. This process is repeated until the student receives at least T_i points, but not more than N_i times. In case of N_i unsuccessful attempts the exercise is suspended and a standard text is given, which recommends learning of the respective subject matter. If the subject areas are not closely related the drill and practice may carry on after registration of poor perfomance in the respective area. Through this simple mechanism (both realization and teacher's use) the system adapts itself to the individual peculiarities of the student.

Another attempt of adaptivity is to react to certain boredom of the students. This is especially important in case of children.

It is applied when a large number n_i of some area A_i (for example above 20) are available. Test items of equal difficulty are generated (In our systems difficulty range from 1 to 9, but practically only 1,2 and 3 are used). Suppose the result of the first third items corresponds to the result between T_i and M_i (i.d. result S_{i1} : $T_i/3 \leq S_{i1} \leq M_i/3$ is reached) and the results of the next sixth of elements falls down considerably (i.d. $S_{i1} - 2S_{i2} = S_{12}$, where S_{12} is a positive number). Then we could presume, that result deterioration comes from certain boredom and fatigue. In this case the drill and practice can be temporarily suspended and some music or animation can be offerred by the computer. The questions concerning the proper value of S_{12} and its parameters (regarding S_{12} as a function) are still open. A large number of experiments are necessary in order to establish if the selected values 1/3 and 1/6 are relevant and in general how efficient this approach is.

If boredom and fatigue, there might be another way of reaction to the high intermediate results. After reaching high number of points in the first third of items, the system generates the second third of items with higher level of difficulty. In this case the treshold number is automatically raised. The new treshold number serves as a criterion if the difficulty level of the last third of items should be further raised.

Another alternative of the explained procedure of UTRAIN may be applied in training students to obtain automatic and quick reactions in some specific areas. In this case instead of raising the difficulty level , time limit for answering each test item and in respect to the whole exircise is introduced.

So far as the examination is concerned, the authors are trying to realize in UTEST the following idea (which requires however more complicated structure of the data base in respect to the subject areas) : subject areas are not regarded as hierarchy, but as semantic net knots. This is a convenient way to express more adequately two, close in respect to the subject, areas (subject proximity). The authors aim by means of this net and the notion of subject proximity at establishing whether certain poor result is casual. It is proceeded in the following way : let S_i ($S_i < T_i$) be a poor result in A_i. Then S_i is registered, but several items from subject proximity areas of A_i are generated (according to the net). Poor perfomance in these close areas means unsatisfactory preparation in general. The disadvantage of this approach consists in overcharging the student. A palliative in carrying out this additional examination might be in doing it after the basic examination.

The authors regard their attempts to create an adaptive systems as in initial stage, at least because sufficient statistical information is not available and because no cooperation with psychologists and pedagogues is yet established. They hope to overcome these difficulties.

FUTURE PROBLEMS

Perhaps the basic problem of the future in computer-aided testing and training turns out to be the so far unsolved problem of computer-aided test item construction. This problem has attracted some methods for its partial solution (9) , (10) , but has not been completely solved yet. For its complete solution are needed most powerful means of the artificial intelligence.

An unambiguous answer should be given to the question to what extent the newest possibilities offerred by the micro-computers - motion, sound etc. can be used in the test items.

The restricions, established by computers and informatics exist practically no more. There exist almost no restrictions concerning the display of formulae, built of symbols of practically arbitrary alphabet, of sophisticated colour illustrations, accompanied even by motion and sound. For the first time it proves out, that not computer technology, but psychology and pedagogy are lingering behind.

Also, particular attention should be paid to the development of new and efficient adaptive strategies.

All these problems may be successfully solved, only if appropriate scientific cooperation between computer scientists, psychologists, pedagogues, linguists is established.

REFERENCES

(1) Ebel R. - Essentials of educational measurement, Prentice-Hall Inc., 1972

(2) Popham J. - Modern educational measurement, Prentice-Hall Inc.,1981

(3) Pavlov R., Eskenasi, Mitkov R. - An adaptive system of children examination and training, Proceedings of the international conference and exhibition on children in an information age : tomorrow's problems today, Varna, 1985

(4) Pavlov R., Eskenasi A., Mitkov R. - KATEST - testing system for administration of the motor vehicle driver's exam, Proceedings of Perscomp, Sofia, 1985 (in bulgarian)

(5) Mitkov R., Pavlov R., Eskenasi A. - Testing in Bulgaria with microcomputers, to be published

(6) Mitkov R. - Computerization of testing - situation and perspectives ,(to be published in bulgarian)

(7) Eskenasi A. - Application No.1 , Contract between the Institute of mathematics and the Sofia Council for popular education, Sofia, 1986 (unpublished)

((8) Lippey G.,(Ed.) - Computer-assisted test construction, Educational Technology Publications,USA,1974

(9) Vickers F. - Computer generated problem sets. A practical approach to computers in education, Educational Technology, 1973, 13(10), p.47-50

(10)Eskenasi A., Sabev V. - A method for a computer-assisted test construction, Serdica Bulgaricae mathematicae publicationes. Vol.11, 1985, p.54-58

(11) Hambleton R., Murray L. - Testing in the United States with computers, in Schlegel J.(Ed) : Bulletin of the inernational test commission and of the division of psychological assessment of the IAAP, No 20,1984