RM-54-11

RESEARCH MEMORANDUM

POSSIBLE USE OF ELECTRONIC COMPUTERS IN MENTAL TESTING

Paper presented at International Congress of Psychology,
Montreal, Canada, June 8, 1954

Ledyard R Tucker

Educational Testing Service

Princeton, New Jersey

July 19, 1954

Session on: Multivariate Methods for Electronic
Computes
Tuesday morning
Room 250, Biology Bldg.

UTILISATION POSSIBLE DES MACHINES A CALCULER ELECTRONIQUES DANS LES EXAMENS PSYCHOMATIQUES

Les machines à calculer électroniques n'ont pas encore été beaucoup utilisées dans le domaine des tests mentaux; mais les nombreux progrès que connaissent présentement ces machines laissent prévoir qu'elles s'appliqueront bientôt à de nouveaux domaines. Dans un avenir encore plus éloigné il semble bien que les automates électroniques serviront de plus en plus, non seulement au traitement des données les plus usuelles, mais aussi aux calculs exigés par les méthodes d'analyse plus complexes auxquelles sont soumis les résultats des tests mentaux.

L'usage actuel des machines électroniques dans les tests mentaux s'est limité à des analyses expérimentales. Une des premières réalisations à signaler fut la constitution d'un Rotateur matriciel: spécialement désigné pour le Secteur des Recherches sur le personnel (Bureau de l'Adjudang général de l'Armée americaine), cet appareil servait à faciliter la rotation graphique des axes en analyse factorielle. Des machines à calculer électroniques, d'un type plus général, ont servi à déterminer les vecteurs caractéristiques at les racines de plusieurs matrices symétriques, dans des études basées sur l'analyse factorielle. Frederic Lord utilisa la machine Whirlwind pour faire une analyse factorielle de 33 tests en 10 facteurs par la méthode de plus grande probabilité, originalement devisée par D. N. Lawley. David Saunders et Charles Wrigley ont aussi utilisé la machine Illiac comme méthode objective de rotation des axes en analyse factorielle. Enfin, dans une étude dirigée par le présent auteur, afin d'établir une échelle de niveau de rendement en vocabulaire, une grande partie des traitements statistiques s'est faite au moyen d'une machine capable d'opérer des calculs sur des cartes.

La construction de nouvelles machines à corriger constitue un progrès remarquable dans le domaine des tests mentaux. On annonce la mise au point de deux machines à corriger susceptibles d'applications très vastes. Elmer J. Hankes a construit une machine capable de faire une lecture photoélectrique des feuilles de réponses et, par voie électronique, de traduire ces informations en résultats et de les imprimer.

Une autre machine considérable est en voie de construction pour E. F. Lindquist. Cette machine utilisera aussi la lecture photoélectrique des feuilles de réponses, de même que la transformation des données et l'impression des résultats par voie électronique. De plus, cette machine pourra servir à cumuler les sommes des résultats et les sommes des produits des résultats obtenus à deux tests par un échantillon de sujets. On prévoit que la mise au point de telles machines aura pour effet de transformer matériellement le travail exigé par des examens psychométriques de grande envergure.

Le contraste vaudrait d'être signalé entre les méthodes d'analyse actuellement employées dans les tests mentaux et celles qui pourront s'élaborer avec l'emploi des machines à calculer électroniques. L'usage de ces machines peut faire disparaître bien des restrictions qui limitent nos méthodes actuelles d'analyse et introduit en même temps des procédés compliqués susceptibles de révéler plus clairement, dans les examens psychométriques, certains effets difficiles à mettre en lumière.

Ledyard R Tucker
Université de Princeton

Translated by: Adrien Pinard

POSSIBLE USE OF ELECTRONIC COMPUTERS IN MENTAL TESTING*

Ledyard R Tucker

Princeton University

and

Educational Testing Service

June 8, 1954

It is a pleasure to me to be able to appear here and discuss with you some of the ways in which electronic computers are impinging on mental testing activities. In these remarks I wish to emphasize and critically examine the aid that these devices may give to our present activities and to those activities that we may desire to undertake in the future. We should be completely clear from the start, however, that these electronic automata are only facilitating dvvices to enable us to perform some important psychological tasks and are not ends in themselves. In this sense I have taken the liberty to interpret the title of this symposium in reverse as "electronic computers for multivariate methods." The major question to which I will address myself is, then: what multivariate methods are there in mental testing whose operations may be facilitated by electronic automata?

In a sense, the giving of a mental test to a person and obtaining a score is usually a multivariate method for evaluating that person in terms of the function represented by the test and the score. Almost all tests are composed of a number of unit variables we call items. An initial problem is

*A paper presented at the International Congress of Psychology, Montreal, Canada, as a part of the symposium: Multivariate Methods for Electronic Computers.

Work of the author summarized in this paper has been jointly supported in part by Princeton University and the Office of Naval Research under contract N6 onr 270-20 and in part by the Educational Testing Service.

that of selecting the items and designating the manner in which they are to be scored. I wish to postpone comments on this problem until after consideration of a second problem, the operations involved in combining the item scores for examinees to their test scores. The wide application of mental testing has made the scoring problem of immense practical importance. Further problems I will consider are those concerned with derived scales for the test scores and with development of other information important to the interpretation of test scores.

Turning first to the scoring operations, let us consider as an initial case the scoring of objective type tests. We will further restrict ourselves to the case in which there are separate answer sheets with distinct spaces to be marked by the examinees to indicate their choices among the given alternative answers. The scoring operation is usually conceived of as the counting of marks in selected spaces. A count of the right answers on an answer sheet will yield a "rights" score for the examinee. If the wrong answers are also counted, a formula score may be obtained as a linear combination of the right answer count and the wrong answer count.

The International Business Machines Corporation scoring machine which has been in use for approximately twenty years was designed to perform exactly these processes. There is wide recognition, however, that this machine is slower and yields a higher proportion of errors than is desired. IBM has embarked on a project to improve its machine.

We would like a scoring machine to be able to perform other operations than the simple counts of rights and wrongs. Several separate counts over selected sub-sets of items may be desired as in a number of the personality and interest schedules. The IBM machine does make counts for three part scores, but the item response positions have to be located in spatially distinct fields. It would be desirable for the items in the sub-sets to be possibly intermixed. Three such part scores are sometimes inadequate. Several special machines for scoring the Strong Vocational Interest Blank have been built and operated by Elmer J. Hankes of the Testscor company in Minneapolis, Minnesota. In this case, about twenty scores are obtained for different selections of items.

Conversion of raw scores to standardized scores would also be desirable. We might wish, also, to combine part scores, using linear weights, to obtain predictive indices.

It would be desirable, also, that the resulting scores would be printed either on the answer sheets or on lists. In case lists are prepared, it will be necessary to print information identifying the examinees for whom the scores are given.

Two new machines have been announced which are designed to perform the aforementioned operations. Both of these machines use photo-electric pick up of information from answer sheets and process this information by electronic components. One of these machines, called TUSAC, has been constructed by Mr. Hankes. The second new machine is being constructed under the direction of Professor E. F. Lindquist of the University of Iowa. New answer sheets have been designed for each of these machines so as to have greater item capacity than the IBM answer sheets. Mr. Hankes designed his machine and answer sheet to be compatable with the IBM answer sheet so as to be able to score these later sheets. Both machines are designed to score a number of tests in a battery at one pass of the answer sheets through the machine. Both machines store the correct answer keys and score conversion tables by electrical means. Speed of operation for these machines is well above speed by present scoring procedures.

Professor Lindquist's machine will have a number of additional features when completed. He plans to have his basic machine completed this fall and then plans to make additional units as feasible to perform the additional functions. The basic machine will not only score but will have some capacity to cumulate sums of scores (raw or converted) for groups of answer sheets. This feature can be utilized to obtain the sums of squares of scores for groups; and, to a limited extent, to obtain the sums of products between scores on pairs of tests. Planned future features involve counters for tabulating item counts and frequency distributions of scores as well as an elementary computing unit for multiplications.

A different type approach to the kind of data processing operation involved in test scoring has been made by the United States Bureau of Census. This agency had a special device designed and constructed by the United States Bureau of Standards to read census schedules and record the information on UNIVAC tape. Tabulations are made from this tape on a UNIVAC computer. If an answer sheet reader were devised and connected to a general purpose computer, many operations might be performed. In scoring, the key could be placed in the machine memory and could be utilized to compare with information from each answer sheet. Several alternative scoring techniques would become feasible such as scoring for patterns of responses or the assignment of different weights to the items. Fairly complex analysis of group statistics could be programed.

We have been considering a very restricted case of test scoring. Let us expand our view a little in the treatment of examinations with written answers. Machines are being developed and tried out for such operations as the sorting of mail. If these machines are successful, it may be possible to have the examinees indicate their choices of answers by written letters or numbers. A further step would be the scoring of items with short, definite answers such as arithmetic problems or the spelling of words.

Before leaving the area of test scoring consideration should be given to the position of individual schools and small testing agencies. Large machines, such as the two new ones previously discussed, are very expensive. In order to justify one such machine there need be a very large amount of scoring to be done. Schools and small testing agencies do not have the requisite volume. Two alternatives exist for these institutions: either to send answer sheets to a central scoring agency or to do the scoring by other means. The IBM scoring machine is of moderate size and will be of assistance to a number of institutions. Mr. Hankes has announced the development of an even smaller machine which should help a further number of smaller institutions.

Let us turn now from test scoring to more analytical activities and the consideration of item selection and weighting. This again concerns a field of considerable activity in testing agencies. In the Educational Testing Service alone, over ten thousand items are analyzed each year. The first question concerning any item is its difficulty in terms of the proportion of given groups that choose the correct answer. A second area of questions about an item is its relation to such variables as the test as a whole, other tests, and school grades or other measures of performance. In order to obtain information for revision of items it is useful to find the number and quality of examinees who choose each distractor, or wrong answer. Present item analysis procedures usually involve some selection of the foregoing statistics and nothing much more complex. The extent of these operations, however, does make it profitable to employ special machines or devices such as the graphic item counter attachment to the IBM scoring machine. Punched card equipment has also been used for these operations. It may be profitable to utilize some more effective electronic device, especially if more intensive analyses are undertaken.

There is considerable agreement that study of the interrelations among items would be of value. We could study the extent of item grouping into homogeneous units and conversely the degree of item heterogeneity. Methods of factor analysis or of latent structure analysis might be employed to disentangle the effects of several proficiencies or abilities which are represented in the composite score of present tests. The electronic computers seem to give promise to aid in this direction of investigation.

In predictive tests for some desired performance, item selection has been based on the cruder item statistics and test constructors' judgment guided by regression studies for whole tests. Professor Philip H. DuBois of Washington University has been conducting a study, supported by the United States Air Force, on the use of complete tables of inter-item covariances for selection of items to predict training school success. While he has developed methods for coping with this problem employing normal punched card equipment, the use of a large electronic computer would have been quite advantageous.

A mental testing problem that is of continuing importance is that of the score scale. I have proposed the development of a type of scale relating test performance to level of proficiency in a unidimensional skill. The procedure involves first a check on the homogeneity of the performances of examinees on representative items. Inter-item covariances are employed and the extent to which they are accounted for by dividing the total group of examinees into subgroups on a score over the items is investigated. The inter-item covariances within subgroups should be approximately zero for a homogeneous set of items. A mean within subgroup covariance over all subgroups is determined for each pair of items and the table of these coefficients is inspected for size and any pattern of relation between groups of items. For a group of items that pass the homogeneity check, the possibility is investigated whether scale values can be assigned to each subgroup of examinees so that the proportion of the subgroups who give the correct answer to each item follows the integral of a normal curve. In case these item characteristic curves can be made simultaneously for a number of items to follow normal curve integrals, a useful score scale would be produced. A trial study is in progress involving 230 vocabulary items given to approximately 6000 students from fourth grade through college sophomores in the United States. Much of the computations have been accomplished on IBM card programed calculators. In case many more studies of this nature were to be attempted, the use of a large electronic computer would be recommended.

Factor analysis has played a major role in mental test developments. Problems of effective procedures, however, have been only partially solved so that methods in common use are to be considered as such approximations as can be performed with desk calculators. More effective methods have been developed such as the maximum likelihood method by D. N. Lawley. Computations involved in this method are complex and can be contemplated only with the use of powerful computing aids. The Whirlwind computer at the Massachusetts Institute of Technology was used by Dr. Frederic Lord of the Educational Testing Service for the analysis of a 33 variable battery to ten factors. Several others have used other computers to obtain principal axes factors from tables of intercorrelations.

Several developments in the rotation of axes in factor analysis are of interest to report. A special machine has been constructed for the Personnel Research Branch of the Adjutant General's Office, United States Army. This machine which computes rotated factor loadings for two factors and displays a plot between these two sets of loadings on an oscilloscope has saved a number of investigators considerable time. The second development is of an analytical rotational scheme by Dr. David Saunders of the Educational Testing Service and Dr. Charles Wrigley, who spoke earlier in this symposium. This method has been successfully tried out on the ILLIAC computer.

Let us consider some of the drawbacks in use of large electronic computers for analyses by psychologists. A major hurdle is the preparation of programs for these machines. Program writing is a complex task requiring special training and skills frequently not possessed by psychologists. An alternative is the use of general library programs for solution of selected types of problems. This definitely limits the variety of problems that the psychologist may consider for assistance from one of these machines. A second major problem is the availability of a machine for use by psychologists. These machines are expensive so that individual psychologists will have to take their problems to centralized computing centers. Fortunately these centers are expanding in number and in extent of services offered.

In taking a look to the future we might contrast operational procedures and methods of analysis now in use with those that might be developed for use with electronic automata. In order to accomplish those things which we are now doing we have learned to accept limitations imposed by facilities available. With more powerful facitities these limits must be revaluated.

We may be allowed to let our imaginations extend to considerably more complex procedures and studies with the hope of being able to realize their accomplishment. Methods of analysis will be less limiting on our gaining of information. than the gathering of observations. This is indeed the ideal, for the importance of observations cannot be understressed; analysis should be effective but consume as little effort as feasible. We may hope in this way to increase the breadth and precision in our knowledge of human behavior.