THE IOWA TESTING PROGRAMS
A RETROSPECTIVE VIEW

E. F. LINDQUIST

University of Iowa
Dr. Lindquist views a scale model of the new Lindquist Center for Measurement, soon to be built on the University of Iowa campus.

The national headquarters of the American College Testing Program in Iowa City, completed in 1968.
The original scoring machine, completed in 1955—a vacuum tube machine, whose output (scores, composites, and names) was printed on continuous paper forms at 4,000 answer sheets per hour.

Here Dr. Lindquist is shown looking at the latest sheet feeder—his own invention—which feeds answer sheets each on computer demand at rates up to 40,000 per hour.

The contemporary model—completely transistorized—transcribes item as well as score data to magnetic tape, which is later processed and printed out by an independent computer.
This short history of the Iowa Testing Programs (ITP) — written for this issue of Education at the insistence of its editors — may perhaps best be introduced by indicating briefly in advance why these programs may be regarded as of any significance outside of Iowa. During the first twenty-five years of their forty-odd year history, these programs grew steadily on a voluntary participation basis until they involved practically the entire school population of the state, as they still do. Throughout this period they constituted an almost ideal “field laboratory” for trying out and gaining experience with new ideas in measurement, the research results of which have been published in many books and periodicals by many individuals. Out of these programs came the Iowa Tests of Basic Skills (ITBS) and the Iowa Test of Educational Development (ITED) which together over the years have been administered over sixty million times to individual pupils in every state, and which must have had some impact on teaching practices and content emphasis. The ITED provided the pattern for and the first forms of the United States Armed Forces Institute Tests of General Educational Development (GED), the present National Merit Scholarship Qualifying Test (NMSQT) and the basic battery used in the American College Testing Program (ACT).

These programs also provided the occasion and the funds for the construction of the first electronic test scoring machine, which with its subsequent improved models has had marked impact on the scope and effectiveness of test usage generally. The services of these machines have been made available by the Measurement Research Center, which has been active in the development of improved methods of test scoring, processing, and reporting that have added to the usefulness and effectiveness of test results. The ITP and MRC, together with Science Research Associates (SRA), made possible the American College Testing Program as well as other state and nationwide services and agencies.

All of this could never have come about had not the University of Iowa and its College of Education been administered during this period by presidents (Walter Jessup, Virgil Hancher, Howard Bowen, and Willard Boyd) and deans (Paul C. Packer, E. T. Peterson and Howard Jones) who through exceptional administration have built a prestige for the University and its College of Education and established an unusually fine relationship with the public schools in which activities like the ITP could thrive. These progressive educators, furthermore, have consistently given maximum support and encouragement and maximum freedom to individual faculty members, resulting in a continuity of key personal and indispensable to the growth of the ITP. In particular, they (together with chief financial officers Fred Ambrose and Elwin Jolliffe) have supported the programs, not by making state or other funds available, but by allowing the programs to retain entirely for their own uses, including research and development, the rather substantial earnings of the ITP. Dean Paul C. Packer’s influence was especially crucial—he initiated the first testing program, encouraged and freed many of his faculty to participate in test construction, and in
general cleared the way for program development—a policy continued by his successors. That the ITP represents very much a team effort cannot be overly emphasized.

The origin of what is now the Iowa Testing Programs was a statewide scholastic competition involving most of the high schools in the state. It was initially called the Iowa Academic Meet—the name echoing the idea of the track meet—and its major purpose was to provide an incentive to scholastic achievement somewhat like that provided in athletics. This academic contest was begun by Professor Thomas Kirby, head of secondary education at the University of Iowa, under the encouragement of Dean Paul C. Packer, and with the cooperation of the Extension Division headed by Dr. Bruce Mahan. The first two years—1929-30—I served as assistant to Professor Thomas Kirby and was responsible for test construction and administration. Then, in 1931, I was given complete charge of the program.

In that program the high schools wishing to participate administered each year a set of approximately one-hour objective achievement tests, each over one of the widely taught standard high school subjects. The tests were administered late in April or early in May. The schools scored their own examinations and reported the results to us. Then, from each school, we selected the top scoring students in each subject. These local winners went to a district contest, where another set of examinations—longer and more difficult—was administered. The winners of the district contests then came to a final contest at the University of Iowa. The newspapers quickly dubbed it the “Brain Derby”!

Every year about a thousand high school students spent two or three days on campus in early June, each taking from one to several two-hour examinations. We had quite a task of scoring and processing those tests overnight, so that we could announce the winners on their last day here. We picked the ten top students in each subject. The final event of this affair was a banquet in the Iowa Memorial Union, at which we awarded medals and some modest scholarships to winning students. It was quite an interesting affair and was, I think, highly regarded by the university administration as a means of attracting some of the more promising students of the state to our campus. Schools as well as individuals competed. The schools having the highest weighted averages on the initial “every-pupil” tests in each of the subjects tested were announced at the final event in Iowa City. Incidentally, schools vied strongly for these honors, and certain schools consistently took top places.

The program had many good features. It did serve well its purpose of arousing greater interest in, and gaining greater recognition for, outstanding academic achievement; it caused the teachers to give more thought to course objectives and course content, and to give more attention to individual students; and it spurred both teachers and students to greater effort. Particularly, it helped to identify talented pupils earlier and more reliably, and demonstrated the unique possibilities of objective tests for this purpose. We found many instances in which students astounded their teachers by scoring at the top of the test distributions—students who before had often been regarded as unpromising or “problem” students and who had regularly been getting grades of C, D and F, perhaps because they refused or failed to follow routine assignments slavishly, and instead did such things as pursuing hobbies or going off to the library to read on their own. Interestingly enough, once those pupils had been identified by the tests as talented, their grades often improved quite suddenly! Startling discoveries of this kind at the high school level are seldom reported today; regular testing in the lower grades now gradually identifies the talented pupil long before he reaches high school.

From the point of view of those of us who constructed and administered the tests at the University, the program served as a golden opportunity to experi-
ment with, and learn how to improve, objective tests — which in the early years of the program were still quite new in education. We could not try out new items in advance, but each year we performed a thorough “post mortem” item analysis on all the tests, and thereby learned a great deal about desirable item characteristics. We worked hard to improve the administrative characteristics of the tests as well, and were, I believe, the first to use separate answer sheets and various improved manual scoring procedures in a wide-scale program.

The Iowa Every-Pupil Achievement Tests (1931-42) were built mainly by department heads in the University High School and other U. of I faculty members, without special compensation. Outstandingly steadfast as authors or editors year after year were: Howard Anderson and George Andrews in the social studies; C. J. Lapp, Alvin Schindler and Paul Kambly in the natural sciences; Harold Lundholm in mathematics; Helen Eddy and Franklin Potter in Latin; and Milton Carpenter in English. Others who assisted repeatedly over the years were John Haefner, Harry Berg, John Briggs, John Partington, Vernon Price, and Ruth Lane. Some of these experts continued to write tests for us even after they had left the campus for positions elsewhere.

In just a few years, our test authors were turning out tests of such good quality as to attract attention outside the state. The Bureau of Educational Research and Service, under Dr. Harry Greene, built up a considerable volume of sales of the tests to non-Iowa users. Ben Wood, Director of the newly formed Cooperative Test Service, became interested in our work, and a considerable number of Cooperative tests were constructed here under my direction.

Beneficial and rewarding as the program was in many respects, too much emphasis on the competitive features soon began to build up. This led to an over-emphasis on the teaching of informational content, and upon the rote learning of facts, since that is what many teachers felt would be most effective in improving test standing. Teachers tended to become more and more teachers of individual subjects, rather than mentors of the pupils most interested in their well-rounded over-all development. In a few instances, teachers were even hired and fired on the basis of the average test standing of their pupils. Furthermore, the striving for awards, per se, was leading to some excessive inter-school rivalries.

We did everything we could, of course, to reduce or counteract these tendencies, short of eliminating the competition entirely. To lessen the competitive aspects, we dropped the district contests, permitted schools to participate in the all-school testing on a non-competitive basis, and renamed the whole program the Iowa Every-Pupil Achievement Testing Program. We tried particularly to emphasize in the tests what we called the reasoned understanding of broad concepts and large units of comprehension, and the ability to do some critical thinking about the materials studied, and we minimized the testing of detailed, specific information. We also tried to draw as much attention as possible, through manuals and otherwise, to these characteristics of the tests, and thus to improve instruction. The tendencies to teach for the tests and to strive for high test scores for their own sake are present to some degree in nearly all applications of standardized achievement tests. The early Iowa program, by accentuating some of these tendencies, gave us special occasion to try to find solutions to a much more general problem. I am sure we learned a great deal about how to build better tests in general as a result of this situation.

Thus, while this program seemed to have undesirable features, it did make clearer to us the tremendous and exciting possibilities of objective tests for educational guidance, individualization of instruction, evaluation of instruction, and other worthwhile objectives. It seemed increasingly urgent that we get away from the subject-matter emphasis
and the competitive aspects, and that we modify the program so as to realize better some of these more promising potentials of objective testing.

Accordingly, by the middle of the thirties, we had worked out a fairly complete plan for what later became the Fall Testing Program for Iowa High Schools. In 1935 I presented this plan at the annual meeting of the Educational Records Bureau Schools in New York City. Several years elapsed, however, before we were able to launch this plan in Iowa — primarily, as I recall, because of the reluctance of the university administration and of the school people throughout the state to give up the very popular “Brain Derby.”

The Iowa Tests of Basic Skills

However, in 1935 we did introduce many of our new ideas at the junior high school level (Grades 6-8), for which we built the first forms of the Iowa Tests of Basic Skills. Here we succeeded from the beginning in placing major emphasis on the development of basic skills and of generalized intellectual abilities, rather than on the rote learning of subject matter, and upon the primary use of tests in the individualization of instruction and guidance.

For this new program we devised grade-equivalent scales, comparable from test to test, and reported results in GE terms. For the first time, Iowa norms of both pupil and school achievement were available to Iowa elementary schools. The statistical services, pupil profile charts, and plotted confidential reports of school averages, were also innovations at the elementary school level.

The immediate success of this Basic Skills Program at the 6-8 grade level was due in very large part to the wholehearted and highly competent cooperation of influential members of the faculty of the College of Education at Iowa — namely, Professors Herbert Spitzer, Ernest Horn, Maude McBroom, and Harry Greene. Particular credit is due Professor Spitzer, then Principal of the University Elementary School, who shared major responsibility for the organization of the tests and for editorial work on the test materials.

In 1940 the Basic Skills Program was extended to Grades 3-5, and additional item writers were recruited to assist in the task of item writing for the annual forms needed. (At this time, too, the Houghton Mifflin company undertook nationwide publication and distribution of the ITBS.) In the editions published from 1940 through 1954 the tests were organized into two batteries — one for Grades 3-5, the other for Grades 6-8. In these batteries, as was true of other achievement test batteries then available, pupils in Grade 5, for example, were required to work through some items intended primarily for Grade 3, and on the other hand, pupils in Grade 3 were required to attempt items appropriate primarily for Grade 5. At no grade level was the pupil presented only with items optimum in difficulty, curriculum placement and content for his own level of development. This was quite radically changed in a new version of the tests introduced in 1955.

The Multi-Level Idea in Test Organization

This new version of the ITBS was called the Multi-Level Edition, and is the design presently being used. The first ML forms were designed and constructed with the major assistance of Professor Albert N. Hieronymus, who since 1948 has been Director of the Iowa Basic Skills Testing Program, and who has been wholly responsible for the construction and standardization of all subsequent forms.

In this version of the ITBS, all items for all subtests and all grade levels were printed in a single booklet, but with a different answer sheet for each grade level. Pupils at different grade levels were instructed to begin and stop at different places in each subtest. Thus there was a different test for each grade level, with the tests for different grade levels overlapping in item content, and each pupil was presented only with items appropri-
ate to his own grade level. Since each subtest had the same time limits for all levels, it was possible to administer the tests simultaneously to all pupils in a mixed group representing several different levels of development. Thus some pupils in the group might be responding to Grade 5 items with a fifth grade answer sheet while their next seat neighbors were responding to Grade 7 items, etc.

I shall have more to say later about the possibilities of this type of test organization and administration with reference to individualization of instruction and of testing.

The Fall Testing Program for Iowa High Schools—The Iowa Tests of Educational Development (ITED)

The long awaited opportunity to introduce the needed new type of program at the high school level occurred in 1942, when the general disruptions of the war enabled us to drop the "Brain Derby" with a minimum of protest from the schools. We had spent five years developing the materials and planning procedures for a new high school program, and thus were all set to go when the time seemed opportune.

The new program presented a number of major new features. First, the basic purpose of the program was to facilitate the individualization of instruction and guidance — to help the teacher become more quickly and reliably acquainted with each individual pupil early in the school year, so that instruction and counseling could be better adapted to his individual needs. Accordingly, the ITED were administered at the beginning of the school year, in September, rather than in May. Teachers thus could no longer feel that they were being measured by the tests, and they no longer had a strong incentive to teach for the tests. Their interest, rather, was shifted to using the test results for the benefit of the pupil.

Second, the whole emphasis in the ITED was upon the development of highly generalized intellectual skills and abilities, upon the understanding of broad concepts, of trends and generalizations, upon interpretive abilities and ability to do critical thinking and problem solving — in general, upon the development of those abilities needed by the student to continue his own education. The battery was thus designed to measure and describe the total educational development of the pupil, rather than the abilities of the teacher or pupil to teach or learn subject matter. The subject organization of the curriculum was ignored. No longer were tests given in Latin II, physics, American history, etc.; rather, the individual tests were concerned with interpretive and problem-solving abilities in broad areas, such as the social studies, the humanities, the sciences, quantitative thinking, etc.

Three of the most important tests were called tests of "ability to interpret materials" in the social studies (Test 5), the sciences (Test 6), and literature (Test 7). These were much more than merely reading comprehension tests — their major emphasis was on problem solving, critical thinking, and sensitivity in the reading situation. Tests 5 and 6 did place a strong premium on the student's knowledge, but only indirectly, in the sense that the more the pupil knew about the broad topics discussed in the reading passages, the more background information he had already acquired, and the more thinking he had already done about these very broad topics or problems, the better he would be able to interpret the test passages and thus earn higher scores, but without being unduly penalized for lack of knowledge of specific, detailed facts.

Incidentally, these particular tests in the ITED have been criticized on the grounds that their intercorrelations are high, approaching the magnitude of their reliabilities, and the critics have therefore contended that the tests measure essentially the same thing, and that a single test would serve the purpose of all three. Some of these critics, I suspect,
would contend that one can quite adequately describe any achievement test’s validity or quality in terms of correlations with external criteria and with other tests, and that one really doesn’t have to look at a test to tell if it is good, granting that one has enough statistical information about it. Such thinking may be somewhat (but not wholly) defensible in dealing with tests intended to predict certain criteria indirectly by measuring something admittedly different from the thing predicted, but it is quite indefensible in relation to achievement tests. To me, by far the most important consideration in achievement testing is that the test items themselves constitute a meaningful and satisfactory definition (in operational terms) of the objectives being measured. If a test is the best achievement test of its kind, this is equivalent to saying that it is its own best available criterion, and relatively low correlations with other criteria and with other measures may only imply weaknesses or defects in those other criteria or measures. Selection or identification of the best of several presumably interchangeable achievement tests can only be done through the exercise of judgment, and never conclusively on the basis of statistical indices alone.

The answer to the criticisms of the three tests of interpretive abilities in the ITED can be satisfactorily resolved only by making a careful inspection and logical analysis of the tests themselves and by deciding for oneself whether or not the abilities or objectives defined by the items in the different tests are indeed intrinsically different. If one is convinced that the abilities (as defined by the items alone) needed to interpret materials and to do problem solving in the field of the social studies are intrinsically the same as those needed in the natural sciences — if one can contend that the teacher of the social studies can just as well accept Test 6 or 7 as Test 5 as a definition of desired outcomes of social science instruction — then the aforementioned critics are right. If these instruments are not interchangeable to teachers of the social studies or the natural sciences, either as definition of objectives or as measures of their attainment, then the separate tests are needed, even though the intercorrelations are high. To me, to contend that because of the high intercorrelations one should substitute a single test for all three is analogous to contending that an optometrist should make all his observations on one eye (right or left) only, or should inextricably merge or average observations on both eyes into a single index, and should then prescribe the same lenses for both eyes — all because of the well established fact that for most heterogeneous human populations right and left eye characteristics are very highly correlated in relation to the reliabilities of obtained measures of them.

The third major feature of the ITED was that the individual tests in the battery were standardized and scaled so as to yield highly comparable scores, expressed in units approximating one probable error of measurement. Thus, the interpretation of the simple graphic profile report for each pupil was made as easy as possible, and the giving by teachers of undue weight to minor raw score differences was avoided. The growth in the student’s development could also be readily estimated by comparing the profiles from year to year. Norms for school averages were provided for each of four categories of school size, as well as for all schools combined.

Fourth, test scoring and all other tasks of a clerical nature involved in the program were performed centrally in Iowa City. (A negative aspect of earlier programs was that they had placed a very onerous clerical burden on the teachers, with an undesirable effect upon teacher attitude toward testing.) Each participating school was provided each September with the necessary supply of tests, manuals, answer sheets, etc., and immediately after test administration the school returned all tests and marked answer sheets to Iowa City. Complete and readily interpretable reports of re-
Results were prepared and mailed to the schools within a couple of weeks after test administration. All of the program services, including provision of tests, manuals, and other materials, and the scoring and reporting services, were provided (and still are) at a very low cost per pupil.

A fifth new feature of the program, perhaps of minor direct interest to the schools but of major interest to the test builders, was the requirement that each participating school devote 20 minutes in each annual program to the tryout (under program conditions) of new test items for future tests. These items are organized into 20-minute self-administering units, which are simultaneously administered, one to a pupil. The units are distributed in rotation to the pupils so that in even a large administrative group each pupil may take a different unit from every other. Each unit, however, contains a homogeneous set of items, corresponding to a single one of the tests, so that the score on the corresponding test in the complete finished battery provides the external criterion needed for analysis of the items. This is far better than the usual ad hoc tryout procedures in which the same tryout items are administered to all pupils in an independently selected sample and the total score on the tryout unit itself is the internal criterion for analysis of items in that unit. From 300 to 400 pupils are used in each tryout sample, and as a result of the rotational system of distribution, practically every school is proportionally represented in each sample. This tryout arrangement provides facilities for the tryout of many more items than are needed in the program itself. I would estimate that, with a total test population of over 185,000 pupils, as many as 7,000 items could, if desired, be tried out under these nearly ideal tryout conditions in a single year. This same provision for item tryout is built into the elementary school (ITBS) program as well.

It is difficult to convey an adequate impression of the importance of these tryout facilities, not only to the Iowa programs and tests, but, as I shall point out later, to other tests and agencies also. Few people who have not themselves been involved realize how tremendously difficult is the task of maintaining test batteries like the ITED and the ITBS — that is, of periodically constructing and standardizing the new forms needed, of writing, trying out, selecting, revising, and assembling the hundreds of items needed for each form, from an initial supply of several times the number used. What is particularly difficult, for most test authors and agencies, is to enlist the cooperation of schools in the special tryout of new materials, to make sure that the tryout samples are representative of the populations for which the test is intended, and to make sure that the tryout items are administered under conditions similar to those under which the tests will be used, or to make sure that the cooperating teachers and pupils will not regard the tryouts as "experimental" nuisances, of no real value or interest to them, and as having no effect on their records. Few people realize, also, how absolutely essential good tryouts are in test construction, or how difficult, if not impossible, it is even for the most experienced item writers to judge the difficulty and other characteristics of items and of individual item responses. It is a matter of more than mere comfort to the authors of the Iowa tests to know that, routinely and without special arrangements, all items will be tried out under program conditions on highly representative and strictly comparable samples, each pupil taking the same attitude toward the tryout items as toward any other items in the regular battery.

As was true of the ITBS, the early forms of the ITED were prepared by or with the help of various U. of I. faculty members and graduate students, including Paul Blommers, Robert Ebel, John Gerber, J. W. Maucker, and Kenneth Vaughn, most of whom who contributed their time and talents with little or no compensation. Rolland Ray not only
constructed several tests in the ITED and ITBS batteries, but also has filled various important administrative and professional positions in ITP, MRC and ACT over a twenty-five year span. Particular credit is also due Julia Peterson, who authored several tests in both the ITED and ITBS series and who has served as my executive and editorial assistant on both programs since their inception. For later forms of the ITED (again as for ITBS) most of the item construction has been done by paid specialists recruited on a nationwide scale, with the major editing being done by myself and my staff. The last four forms of the ITED, including the two issued this year, were built under the direction and editorship of Dr. Leonard Feldt, who has directed the Fall Testing Program since 1957.

Building a Scoring Machine

We had hardly begun the ITED program when a young publishing house in Chicago, Science Research Associates, became keenly interested in it and asked for an opportunity to publish the ITED outside of Iowa. It seemed desirable to keep the nationwide use of the ITED similar to their use in Iowa — that is, to have the publisher sell a complete package, including the tests, scoring services, reports, interpretive materials, and so on, rather than just sell test materials. We felt that the tests could be much better protected, and that better practices would result, if the schools were not to buy or keep on hand any materials at all, but were simply to rent the tests and receive a complete service with them. This was regarded as a rather radical innovation in test publishing at the time, and SRA seemed more interested than other publishers in trying this new plan. So we arranged that SRA would take over the nationwide publication and promotion of the ITED battery and services, while we took on in Iowa City the task of scoring and processing, not only the ITED tests given in Iowa, but those given outside of Iowa through SRA as well.

The high school program immediately gave rise to a demand from the elementary school people for similar scoring and processing services on the Iowa Tests of Basic Skills. We were not able to satisfy this demand at once, since we lacked the facilities to do the volume of scoring involved. In the late forties and early fifties we no longer had on campus the large number of returning war veterans whose wives had earlier constituted a major source of good clerical help, and we were finding it difficult to hire enough people to score the tests for the high school program alone. Certainly we were not in position to take on any extra scoring load for the elementary schools.

It seemed apparent then that we simply had to have some better way of processing test materials. There had been earlier attempts to build scoring machines, of course. IBM had developed one (the 850), which we had tried in the “Brain Derby” program and in the University Examinations Service, but it seemed to us inadequate for large-scale processing. I had been accumulating ideas concerning machine scoring on my own for quite a long time. Having majored in mathematics and physics in college and having begun my graduate work in mathematical physics at the University of Chicago, I did have a little background for tackling this problem. Fortunately, at just the right time (about 1950) I consulted on this problem with a good friend of mine, Professor Phillip Rulon at Harvard, who encouraged me to build an electronic (rather than electrical-mechanical) machine and worked with me on the initial design. This was a fairly new idea in those days, the first commercial computer having just been built, and the venture presented more difficulties than we had anticipated.

The trouble was, really, that the state of the electronics art was not quite yet ready for the kind of machine we were trying to build. For instance, the original design was based on photo-sensitive devices of a vacuum tube type which were too bulky to be readily brought to bear on the closely spaced marks needed on
the answer sheets, and which proved far less stable than our early tests of them had promised. It was not until machine construction was two years under way that a compact solid-state photo diode became available and one which happened to have exactly the right dimensions and electrical characteristics required! Because of what seemed for a while to be unsurmountable technical difficulties, we were several times just on the verge of abandoning the entire venture, and would undoubtedly have done so had it not been for the ingenuity and loyal stick-to-it-iveness of MRC's chief engineer Robert Edberg and his associate Andrew Veranais and other helpers.

Needless to say, after having made what many regarded as impossible claims for our machine in a presentation at the ETS Invitational Conference in New York in the fall of 1953, I was greatly relieved when new technical devices and the persistence of our engineers finally saved the day! For the construction of later models, we have been fortunate in having the services of an electronics engineer in the genius class — John McMillin, who, together with unusually competent associates and technicians, especially George Carsner and Richard George, has produced many remarkable improvements over the earliest model.

Measurement Research Center

In 1953 we founded Measurement Research Center. When we were pretty well along with the design and manufacture of this new electronic scoring machine, we realized that, once it was built, it would have a capacity far beyond the needs of the Iowa Testing Programs. We felt almost obliged to make its services available to schools throughout the country, in the hope of relieving teachers everywhere of test-scoring burdens. But we also realized that to set up a high-volume scoring service here on the campus, under the auspices of the University, would not be an appropriate university activity, since we would be competing directly with commercial agencies of various kinds. So we decided to set up an independent, non-stock, not-for-profit corporation, which would operate the machine and provide a variety of scoring and processing services. The members of the Board of Trustees were key administrative officials of the University. The trustees and officials of the Corporation, including myself as president, served without salary. The corporation charter provided that any surplus income over expenses could only be spent on educational research at the University of Iowa.

The original scoring machine scored at the rate of over 4,000 sheets per hour, converted raw to standard scores, computed composites, "read" the pupil's (coded) name from the sheet, and printed out a paper record of names and scores. We soon realized that what we needed was a machine that would feed raw scores and item data directly into a computer, which could then perform a wide variety of analytical and reducing operations as the scanning proceeded, and which could employ the usual tape, card, and paper output devices of the computer. To this end, we designed and built a series of consecutively improved scanners, culminating in the present "Model 12," which scans sheets at the rate of over 40,000 per hour and is capable (with its on-line computer) of performing virtually any operations on the data and preparing any types of reports that are of practical interest. I was deeply involved in the development of these machines, and personally designed, engineered through its early stages, and patented the high-speed paper feeder which is now employed in Model 12, and which in the prototype model fed sheets and cards at rates up to 100,000 per hour. Incidentally, it is the computer, rather than the mechanical (paper feeder) component, that sets the present 40,000 sheet per hour speed limit, even though MRC is using one of the fastest computers in the business.

One of the more important developments in MRC processing was the perfection of a system for scoring stapled test booklets (up to 48 pages) by cutting off the stapled fold in an automatic machine.
of my own invention, thus separating the booklet into separate sheets, and then scanning the pages consecutively to produce an integrated report for the entire booklet. This development is having a marked effect on testing procedures at the lower grade levels and for disadvantaged children who are not capable of following complicated directions for marking separate answer sheets but are capable of the simpler marking of the responses themselves on the test pages. I expect this development to become of major significance, since it will eliminate the restrictive influences of the separate answer sheet on test construction, and will permit the construction of improved “mark the booklet” types of tests for all pupils at all levels.*

Other Tests and Agencies Stemming from the Iowa Tests

While the ITBS and the ITED together have long been given each year to many millions of elementary and secondary school pupils throughout the United States, and have thus influenced educational practices directly, the Iowa Programs may also be regarded as of some significance because the tests and the experiences gained with them have facilitated or contributed to the development of other major tests and test agencies. The most important of these are, I believe, the USAFI Tests of General Educational Development, (USAFI GED), the Qualifying Test (NMSQT) for the National Merit Scholarship Program, and the basic battery used in the American College Testing Program (ACT).

The USAFI Tests of General Educational Development

Early in the participation of the U.S. in World War II, it became apparent that American education faced major problems of providing for the continued education of personnel in the armed services while they were yet in service, of evaluating and accrediting these and other informal and incidental in-service educational experiences, and of facilitating and regulating the return to school of service men and women whose formal education (particularly at the high school level) had been interrupted by the war.

The United States Armed Forces Institute was organized by the joint services to help solve these problems, and a Civilian Advisory Committee was appointed to provide policy leadership. Among many other things, this committee saw the need for a test battery, to be administered to service personnel near the time of their discharge, that would evaluate their educational development at that time and thus provide a basis for certifying if they had achieved the equivalent of a high school education and were ready for admission to college without having to return to high school. The need was for a test that would measure as directly as possible the individual's ability to perform the kinds of tasks, solve the kinds of problems, do the kinds of thinking required for success at the college level, without penalizing him for not having acquired some of the unique informational content of standard high school courses.

The then new ITED seemed to me to have exactly the qualities required, and I found ready acceptance of this suggestion by the Advisory Committee, of which I was a member. The result was that I was asked to take charge of the construction and standardization of what were to be known as the USAFI Tests of General Educational Development (GED). Fortunately, we were able to introduce the new program immediately by constituting the first form of the USAFI GED out of five of the tests in the new ITED, and were able to establish standards for this battery almost immediately through the splendid cooperation of the high schools of the country in a nationwide standardization program. For two years I worked with Ralph Tyler's USAFI Examinations staff at the University of Chicago, directing the building of subsequent high school and college forms more specifically adapted than the ITED to service needs.
Since then, many additional forms of the GED have been built by other agencies, and the GED tests have been administered to millions of service and ex-service personnel, as well as to many (non-service) workers in industry and government.

The Qualifying Test for the NMSC

About 1957, John Stalnaker, President of the National Merit Scholarship Corporation, saw the possibilities of tests of the ITED and GED type for the selection of college-bound students meriting scholarship aid. He recognized also the unusual facilities of MRC for providing the prompt and accurate scoring and processing service needed in the National Merit Program. Tests measuring directly the student’s ability to perform the complex intellectual tasks needed for success in college seemed more likely to provide dependable and valid predictions of college success than tests of abstract skills which could be validated only indirectly by correlation with the somewhat suspect course-mark or grade-point-average criterion of college success. Stalnaker recognized also and valued highly the superior facilities of the ITP for tryout of the masses of items needed for the many new ongoing forms of the NMSQT. Accordingly, in 1958, SRA took over the Qualifying Test Program for the NMSC, and in turn contracted with ITP and MRC for the test construction and processing. As with the GED, the first new forms of the NMSQT were “lifted” from the ITED. The NMSQT is now constructed, scored and processed by SRA, using scoring machines built for them by MRC, but for many years all items used in the NMSQT have been tried out through the facilities of the Iowa programs.

The American College Testing Program

In many respects, one of the most important of the “spin-off’s” of the Iowa Testing Programs was the American College Testing Program. This program, launched in 1959, was designed to serve the needs of the large state universities, state and municipal colleges and junior colleges, and the great majority of the smaller private and denominational colleges of the country — generally, the institutions that were not already served by the College Entrance Examination Board. Instead of having a problem of selective admission, these institutions were anxious to increase their enrollments, or tried as a matter of policy to increase their facilities as needed to provide for all qualified high school graduates who sought admission. They did have a recognized need for pre-admission tests, but primarily to provide information needed in advising the students at the time of first registration, and for general guidance and counseling. A very small number of institutions had tried to administer their own tests for these purposes, but with unsatisfactory results. Testing during freshman week presented too many difficulties in processing and reporting the results in time for use at registration, while tests administered in high schools on a statewide basis failed to reach not only out-of-state students but many others as well. Furthermore, practically every institution lacked the facilities to develop by itself each year the new and secure forms of the tests that were really needed.

Following World War II and Sputnik the surge of students to the colleges increased the need for pre-admissions testing among these institutions. Their need, however, was not for a test that would skim the cream off the top of a distribution of applicants, but rather for one that would help screen out the few who might least profit from college opportunities. What these institutions really needed was a test battery that would describe the student and define his educational needs and abilities in meaningful differential terms, terms related to different areas of instruction — tests that each institution could use, not only in screening and advising at registration and in later counseling, but also in better defining its own educational objectives, so that it could modify or revise
Education

its curriculum and offerings better to fit the needs of the students who were admitted. Again, the real need was not for an efficient predictor of a single doubtful grade-point-average criterion, but a direct test of the student's general educational development that would itself constitute a criterion — a criterion useful in evaluating the institution's own educational practices. In other words, what appeared to be required was a test of the ITED/GED type. Most important, the advance information needed about students went far beyond test information alone. Indeed, each institution needed to know everything about the student that is significant and relevant to his college needs; his past achievements, his interests, his educational and vocational aspirations and plans, his financial needs, and so forth. Furthermore, lacking its own data processing facilities, the average college needed to have this and the test information organized, analyzed and digested for it with special reference to the unique characteristics of the institution's own peculiar student body, and reported to it in the most readily usable and interpretable form possible.

What was clearly required was a new testing and data collection agency conducted on a nationwide scale. For any such agency to succeed, at least seven things were essential. First, a test battery appropriate to the needs just described — a battery already fully developed and publicly accepted, provided with national norms for a meaningful interpretation of the first test results, and backed by convincing evidence of validity, especially in predicting college grades. Second, a test development facility for the continuing production of the many new and secure forms of the test needed in the ongoing program, including a facility for trying out or field testing annually the very large number of items needed. Third, an experienced publication and distribution facility, for preparing and printing the tests, forms, manuals and promotional materials needed, and for establishing test centers and distributing the materials to them. Fourth, a test scoring and data processing facility, capable of scoring accurately and reporting results promptly for whatever number of students might be tested at any one time. Fifth, the risk capital to pay all developmental, promotional, and operational costs until the program could become self-supporting. Sixth, an institution, agency or corporation that would sponsor or serve as host to the program, and be legally responsible for it; and seventh, a plan for securing the necessary participation and cooperation of the colleges.

I have already described or suggested how most of these necessary ideas, materials, experiences and facilities had been gradually and slowly developed in the ITP and MRC and at SRA over the years preceding the initiation of ACT. During those years, I had served on the standing committee of the College Board for the Scholastic Aptitude Test, and thus had special occasion and opportunity to think about alternatives to the improvements upon the Board's program. During these years also the President of SRA, Lyle Spencer, and I had many discussions about the possibility of utilizing the combined facilities and resources of SRA, ITP and MRC in the college area. In 1957 SRA took over the Qualifying Test Program for the National Merit Scholarship Corporation, using (as I have already indicated) a test derived from the ITED and the facilities of ITP and MRC for this purpose. The immediate success of this venture convinced us that the time was ripe for launching a broader program competitive with the College Board's.

At this point, recognizing that the cooperation of the colleges of my own state and of college admissions officers everywhere was essential, I approached Ted McCarrel with an outline of suggested program services, and with the proposal that he and I jointly launch and direct such a program. McCarrel at that time was Registrar of the University of Iowa, and was very influential among college admissions officers, having just finished a term as president of the
National Association of College Admissions Officers. My proposal provided that the test battery to be used the first year was to consist of four tests adapted from the ITED, along with the norms and the great mass of evidence of validity already available for these tests. MRC at its own risk would serve as the corporate host of the program, would pay the costs of organizational meetings and of the large amount of travel required of Ted McCarrel and myself and the regional co-ordinators, and would also handle the test scoring and data processing for the program, as well as the research service to individual colleges. SRA agreed at its own risk to provide all of the remaining services and materials needed and to supply the necessary risk capital.

McCarrel took up this challenge with real enthusiasm, and his leadership constituted a very major factor in the subsequent success of the program. Later, on March 2, 1959, the Board of Trustees of MRC authorized its officers to establish the American College Testing Program as a division of MRC, and ACT became a legal entity. However, we had felt from the beginning that the ACT Program should be run by, as well as for, the colleges, and on a not-for-profit basis. As a first step in this direction, in 1960 we organized an independent ACT corporation, the members of which were the various state coordinators. MRC, without compensation, then gave up its title to ACT to this corporation.

For the first year, on a part-time basis, McCarrel and I served, respectively, as General Director and Director of Research and Development. In 1960 ACT appointed a full-time president, and McCarrel and I became members of the Board of Trustees. In 1964 I helped disqualify myself for Board membership by presenting a resolution for reorganization which provided that ACT should be governed by a board whose members had no administrative or financial involvement in ACT but served only in the public interest. Since that reorganization was effected, in 1966, I have had nothing to do with the operation of ACT, serving only nominally as Senior Consultant to the Board. From the beginning to the present, all items used in the many forms of the ACT battery have been selected and revised on the basis of tryout data obtained through the facilities of the ITP.

The Iowa Educational Information Center (IEIC)

For many years the ITP annually collected a tremendous amount of test item and score information, as well as other demographic data, about Iowa's school population, but until 1963 no means had been provided for storing, integrating, and making this wealth of information accessible to state and local school administrators and to research workers in education. In late 1963, to meet this need for a centralized data-bank and statewide data collecting agency, we established the Iowa Educational Information Center with the help of one million dollars in grants from the Ford Foundation and the United States Office of Education. This again, was a project that was made feasible only by the unusual data collecting and data processing and computer facilities of MRC. The principal mission of this project, now under the direction of Walter Foley, is to develop information systems for the public schools that will facilitate the kinds of informed decision-making by Iowa school administrators needed in the progressive schools of the seventies.

Other Professional Activities

I have been requested by the editor to include in this narrative some reference to possibly significant activities not necessarily related to the ITP.

Throughout the thirties, forties, and into the fifties, I tried to carry my share of the teaching and research load in the College of Education, offering courses in tests and measurement and statistical methods, and conducting and directing research in measurement. In this connec-
tion, I published *A First Course in Statistics* in 1938. If there was anything innovative or noteworthy in this, it was the effort to introduce the use of the Socratic method in the teaching of statistics, through the use of a workbook published as *Study Manual for a First Course in Statistics*.

Through the thirties, practically all of the significant work being done in the application of analysis of variance was that in the fields of agriculture or agronomics, and nearly all of the related theory and application was presented and discussed in terms meaningful only to students of agriculture and animal husbandry — plots, blocks, yields, treatments, litters, etc. — rather than in terms such as schools, classes, scores, methods, etc., familiar to students of education. This, plus the characteristic "cookbook" presentations employed made the available textbooks unnecessarily difficult for students of education. What was needed, it seemed to me, was a text that not only "translated" these discussions into educational terms, employing models and examples chosen from education and psychology, but that also attempted to develop a critical understanding, with specific reference to education, of the analytical procedures employed and especially of the basic assumptions involved. This need prompted me to write *Statistical Analysis in Educational Research*, which was published in 1940.

Experience with this text convinced me that there was urgent need for a much more comprehensive and advanced treatment of analysis of variance, one that would present many recently developed models and designs useful in educational and psychological research. What seemed needed, most of all, was an entirely new and education-oriented classification or hierarchy of standard designs. In response to this need, in the late forties I devoted two years of especially intensive and sustained effort in writing *Design and Analysis of Research in Education and Psychology*. The book was published in 1953 after two years of classroom trial and revision.

The American Council Handbook on Educational Measurement

During the thirties and forties, an urgent and constantly growing need had developed for a comprehensive handbook and reference on the theory and technique of educational measurement, for use particularly in the advanced training and practice of measurement specialists. A great deal of knowledge and experience concerning the art as well as the technique, theory, and philosophy of measurement had been and was being accumulated, but much of it had never been written up and published or was accessible only in scattered articles in a large number of periodicals.

The Committee on Measurement and Guidance of the American Council on Education had long been aware of this need, and had sponsored an elementary book of this character, *Construction and Use of Achievement Examinations*, edited by Hawkes, Lindquist and Mann, published in 1936. In 1945, the American Council Committee called a special conference to plan and initiate a much more advanced and comprehensive volume. At this conference, a tentative table of contents was prepared, and a total of approximately seventy measurement workers — including practically everyone of prominence in the field — were nominated for participation, twenty as chapter authors and the rest as collaborators, with myself as editor. The plan was quickly approved by the American Council, a supporting grant was obtained from the Grant Foundation through the offices of Ben D. Wood, and work was initiated on the project early in 1946.

The task of eliciting the participation of chapter writers and collaborators, arranging for effective team work, and "riding herd" on the preparation of manuscripts proved to be much greater than anticipated. Four years rather than the expected two elapsed before all of the planned manuscript was in the hands of the editor. The individual chapters, almost without exception, were of very
high quality, but in their original form were not closely coordinated or integrated with one another, and as expected, a number of interstitial topics had not received adequate treatment. Hence, a large amount of rather stringent and rigorous editing seemed called for. In my opinion, the feature of the project that made this kind of editing possible and acceptable to the contributors is that much of it was done cooperatively, on a group basis, by the major contributors themselves. A two-week "work conference" involving eight of the authors of major chapters was held at Gloucester, Mass., in August, 1949. Working in editorial teams of two or three per chapter, these people carefully read, criticized and revised more than 1,100 pages of manuscript, and prepared specific suggestions to the authors for further revision. It was much easier for me later to supervise and encourage the execution of these and other suggestions than it would have been to attempt singlehandedly to bring about the desired changes. In consequence, we were able finally to produce what, for a symposium involving so many individuals, seemed to me a reasonably well integrated, coordinated and inclusive treatment of the theory and technique of educational measurement.

The volume, under the title of *Educational Measurement*, was finally published by the American Council in 1950, almost five years after the planning conference. It has, I hope, exercised an appreciable influence on measurement practices and in the training of measurement workers, not only of nationwide, but also of international, scope.

**The Future of the Iowa Testing Programs**

Upon my own retirement from the university faculty in July, 1969, William E. Coffman, formerly of Educational Testing Service, was appointed Director of the Iowa Testing Programs and Lindquist Professor of Educational Measurement in the College of Education. The administration of the individual state-wide testing programs had long since been delegated to A. N. Hieronymus, Director of the Iowa Basic Skills Program since 1948, and Leonard Feldt, Director of the Fall Testing Program (ITED) since 1957. These programs have continued to grow each year since their inception, until now practically the entire public school enrollment and much of the parochial and private school enrollment of the state is regularly involved on a voluntary participation basis.

When arrangements were first made with Houghton Mifflin Company of Boston to publish the ITBS for sale outside of Iowa, it was agreed that the total royalty should be paid to the University of Iowa in the name of the Iowa Testing Programs. On the multi-level version of the ITBS and on the SRA editions of the ITED, a ten per cent royalty (not including editorial royalties) is likewise paid to the ITP. Over the years, these royalties have constituted a very substantial income to the programs. This income has enabled ITP to support hundreds of research fellowships in the College of Education, to subsidize numerous campus projects and departments, and to pay substantial portions of many faculty salaries. It was this income which enabled ITP to finance the first electronic scoring machine. It has also enabled ITP to acquire a significant reserve, most of which is now to be spent on a $3+ million building to house the ITP, the University Computer Center, the Iowa Educational Information Center, and the Division of Educational Psychology, Measurement and Statistics of the College of Education. This building, paid for entirely out of ITP and MRC earnings, will be known as the Lindquist Center for Measurement.

The question of the future of the Iowa Testing Programs, and of the ITBS and ITED, is of course a part of the larger question of what will be the future role and importance in American education of comprehensive standardized achievement test batteries such as the Iowa tests. The distribution of emphasis over the various uses of such test bat-
I believe, however, that in the ungraded school of the future, even more than in the graded schools, such testing will have to be individualized as well. In the truly ungraded school, the same uniform test battery will be even less appropriate for all pupils even in a single years-of-schooling group, because of the greatly accentuated individual differences that will develop under highly flexible individual scheduling. What will be needed will be a system of testing that is as highly individualized as instruction itself.

The organization of the ITBS anticipates and suggests a solution to this need for completely individualized testing. A single test booklet already contains all items for all levels over a very wide range of development (in present terms, Grades 3 to 8). For each of the eleven skills tested, the time limits and administrative directions and conditions are identical for all levels. At present, the pupils in a mixed grade group can all be tested simultaneously if each pupil is provided with an answer sheet appropriate to his own grade level, but this still permits no variation in level from one
skill to another for the same pupil. To achieve complete individualization, it remains only to provide a single comprehensive answer sheet or folder for all levels combined, corresponding to the complete test booklet. Each pupil could then be provided with typed individual instructions, telling him with which items he should begin and end in each test. All pupils in a highly variable group could then be tested simultaneously, each pupil in a large group conceivably taking a different combination of levels in the eleven tests than every other pupil in the group. In a computer monitored system, the individual instructions would be computer prepared, and the combination of test levels most appropriate to each individual would be selected on the basis of past test performance and other available data in the computer's information bank.

The Iowa Measurement Research Foundation (IMRF)

The most recent event in the chain of ITP-MRC developments is the formation of the Iowa Measurement Research Foundation.

From 1955 to 1968 Measurement Research Center had been growing steadily at a very satisfactory rate, and in 1968 had over 230 permanent employees. Most of these were young persons at a professional or highly skilled level — systems developers, programmers, computer operators, engineers, technicians, accountants, and managers — plus a competent clerk corps. Such an organization must continue to grow if it is to hold young personnel who are in high demand elsewhere. By 1968 there were signs that MRC's test scoring and processing business might begin to level off, partly because an increasing number of large school systems and publishers were acquiring their own scoring machines. The answer seemed to lie in diversification of MRC's activities, which in turn would require a considerable investment of capital at some risk. Not being interested just in the financial success of MRC for its own sake, nor anxious to assume responsibility for the needed expansions and investment, and possibly influenced by my own impending retirement, the MRC Board decided that a better alternative would be to sell MRC to an agency that could bring in the desired new business and that had the necessary risk capital. After considering several possibilities, the Board decided that Westinghouse Learning Corporation (WLC), which was interested in acquiring MRC, offered the best promise of maintaining MRC in Iowa City in a way that would continue to benefit the University and the community.

Accordingly, in July, 1968, the name and all assets of MRC were sold to WLC for five million dollars in Westinghouse Electric Corporation common stock. The Board then changed its corporate name to the Iowa Measurement Research Foundation (IMRF), retaining the same officers, with myself as president, as had served the old MRC. IMRF, under the original charter of MRC, is committed to spend its net income only on educational research at the University of Iowa, and in the event of dissolution of IMRF, its assets become the property of the University. My principal retirement activity during the coming years will be to help the Board apply IMRF's resources to the further betterment of measurement practices and to the encouragement and support of educational research at the University of Iowa.

I have tried, throughout this narrative, to give recognition to the many key individuals whose cooperation, contributions and services have been indispensable to the development of the ITP or related activities. I am sure I have missed many officials and proponents who deserved mention, including literally hundreds of loyal and highly competent employees of ITP and MRC. This whole venture has represented in the most genuine sense a cooperative team effort, in which no single individual deserves the lion's share of the credit.