THE IDENTIFICATION, DEVELOPMENT, AND UTILIZATION OF HUMAN TALENTS

By: John T. Dailey, University of Pittsburgh

For at least the past 25 years, there has been serious discussion of the need for a program of research which would, on a nationwide basis, yield factual information about the nature and distribution of human talents and how best to assist individuals to identify, develop, and use them. There has been an urgent need for a comprehensive survey including aptitude, interest, motivational, and background factors, to be accompanied by intermediate and long-range follow-ups of the individuals studied in the survey. Such a survey, Project Talent, is now underway.

Through the cooperative assistance of the U. S. Office of Education, the National Institute of Mental Health, the National Science Foundation, and the Office of Naval Research, such a research study was planned, and the testing was carried out during the school year 1959-1960.

It was planned that the two-day battery of tests be representative of the best current types of aptitude and achievement tests that have been proven to be useful and important in selection, measurement, and guidance of young people at the secondary school level. There were also questionnaires to obtain important background information regarding the previous experiences of the students: their neighborhood, family, community, interests, activities, etc. There were also measures of their motivations, levels of vocational and economic aspirations, and personal plans. The tests were designed to measure the functions covered by most of the standardized tests now widely used with high school students.

The tests that have been used in the project were prepared especially for this research following detailed specifications developed by the staff and advisory panel following a year's intensive study. No test was an exact parallel of any existing commercial test. The tests will be withheld from subsequent publication and will be reserved for research use only.

In the planning phase of Project Talent, one of the first tasks in developing the test battery was to decide on a number of types of tests to be developed. The Staff were assisted actively in this by the Test Panel and the Executive Committee. In September 1958, following the Test Panel meeting, a number of test types were selected which were believed to cover most of the important variance in aptitude and achievement tests in current use and which were suitable for administration by teachers in public high schools and other secondary schools under the logistic limitations of the Project. In the early spring of 1959, detailed rationales and specifications were prepared by Staff for each of these types of tests. These specifications gave a detailed description for the type of test to be developed in the battery, giving its purpose, what it was supposed to measure, the number of items with their difficulty levels, and the like. After these had been reviewed and approved and following the Test Panel meeting in the spring of 1959, the Staff prepared two forms of each type of test and they were administered in May 1959 to approximately 6000 students in eleven schools. The purpose of this administration was to obtain item analysis data and test inter-relationships data to guide the staff in devising the final form of the Project Talent battery.

One of the most important considerations in the design of the Talent battery was conserving testing time to the utmost. Although ten hours of testing time may seem to be a great deal, when it is necessary to include in this time as many different measures as were strongly recommended by the Panels and the Committee, there is not a minute to spare. Also, for any given test to be included, it must meet the very strong competition of all the other tests very strongly recommended by various individuals within the Panels and Committees. Virtually all the test types recommended also had been widely used and had proven themselves by demonstrating value for use by counselors and research personnel. Accordingly, it was necessary that each test demonstrate empirically that it had important variance to contribute over and above the variance covered by all the other tests in the battery. In all, fifty-four measures were scored and included in the analysis.

It proved not to be possible to accomplish all of the scoring of such a large battery during the limited time available with any known existing equipment. Also, any known existing equipment would have been extremely expensive, since it was necessary to score each test separately for two separately-timed halves and for total score. It was also necessary to score rights and wrongs separately and combine them into formula scores. Further, some of the tests were on more than one answer sheet, and some of them consisted of items scattered throughout a 350-item booklet. The scoring for each individual was equivalent to 750 rights scorings. To score the test conventionally would have meant obtaining these 750 scores and then combining them into the relevant scores, total scores, etc.

The scoring and item analysis were accomplished simultaneously on an IBM 704 computer. A special program for this purpose was developed. This scoring and item analysis program scores all the tests in the battery, computes the half-scores separately plus the total score, is flexible to score by formula (rights minus wrongs, rights minus one-fourth wrongs, etc.), and it computes a point biserial correlation coefficient between each item choice and the test score it appears in. Later, it prints out these results in roster form for sending back to the participating schools. Existing programs were used to compute comparable half reliabilities, inter-correlations, the multiple correlations between each test versus all the other tests in the battery, and the partial correlations between each pair of tests in the

battery with all other tests held constant. The regression program also computes the inverse of the matrix and the regression weights. This program can also use any one variable as the dependent variable and call in each of the other tests successively, one at a time, in the order of its partial correlation with the dependent variable, and compute the multiple correlation and raw score regression weight for each successive combination of tests.

The final step in the analysis of the pre-test data was to compute the uniqueness coefficient which consists of the reliability coefficient minus the square of the multiple correlation coefficient for the test versus all the other tests in the battery. This uniqueness coefficient was very valuable in helping to evaluate the tests in the experimental battery. Most of the tests turned out to have some useful unique variance. This is not surprising since each test tried out was believed by those sponsoring it to have an important unique contribution. After all of the statistical data had been completed and interpreted, several tests were eliminated from the battery.

It has been found that, once the items of a test are on tape, it is about as cheap to score and analyze simultaneously as it is to analyze alone after the tests have been scored in a separate operation. It is believed that this has profound implications for the scoring and analysis of tests. Efforts are under way to follow through and develop a combination scoring and analysis routine, known as Test Assay, to capitalize on these features. Test Assay may be used as a description of the total process of determining the communality inter-relationships and also the uniqueness of any specific test included in a given battery. This, in a sense, assays a test both in terms of its overlap with the other tests tried out simultaneously with it, and in terms of its unique potential contribution. Of course, the uniqueness of a test is specific to the particular combination of tests it is tried out in. This makes it important to use a standard, known, highly comprehensive and non-redundant battery for this purpose. It is believed that the Talent battery should be very useful for this purpose. Fortunately, it is easy and economical to iterate this process by cycling through it several times and changing the combination of tests as indicated.

It is also important that Test Assay be carried out on at least moderately comparable samples since it is well known that reliabilities and correlations vary a great deal from one kind of population to another. However, one of the most interesting aspects of the uniqueness coefficient is that is appears to be highly stable from sample to sample as compared with the stability of reliabilities and correlation coefficients. Apparently, if because of restricted range or some other factor, the reliabilities tend to be lower on one sample than another, the multiple correlations will be correspondingly lower and thus the uniqueness coefficient will tend to be stable.

The key to the feasibility and economy of this approach is the adequacy of the sensing mechanism for putting the answer sheets on tape. In the pre-test tryout, the answer sheets were run through a type of document reader that automatically sensed and punched cards at a price only slightly cheaper than hand-punching. The Measurement Research Center at the University of Iowa now has a document reader punch which will punch one card per answer sheet insertion at a speed of about 6000 per hour. It will get about 120 items on each card by means of punching two items per column. It uses a special answer sheet similar to the regular type multiple choice answer sheet. It is highly flexible and alpha-numerical in character. If desired, test items can be printed directly on the answer sheet. It can also be used as a separate answer sheet.

The next step in the development of the art in this area will be the development of a document reader which can put an entire answer sheet of 500 multiple choice items directly on magnetic tape at a speed of 6000 per hour or greater, and without going through the intermediary of a punched card. This would have the important advantage of eliminating the necessity of dealing with large numbers of cards and having to sort together several cards for each subject. It is reported that developments of this kind are under way by several groups.