IDENTIFYING THE FUTURE SCIENTIST

Marion F. Shaycof
University of Pittsburgh
American Institute for Research

(AAAS Meeting; New York; 30 December 1960)

One goal of Project Talent is to develop ways of identifying
students with exceptional aptitude for science and math very early in their academic careers--early enough that the talent will not be dissipated through ignorance of its existence.

Before going into detail about steps towards this goal I want to reemphasize one point that has already been indicated. Project Talent is not concerned solely with the potential scientist and mathematician. The purposes of the Project are manifold. We are interested in many kinds of aptitudes and many kinds of talent--literary talent, linguistic talent, mechanical talent, etc.

But though the identification of students with unusual aptitude for science or mathematics is not the sole purpose of the project, it is unquestionably a major one. There is probably little need to convince this audience of the importance of being able to determine which students have talent along these lines.

This purpose imposed on all aspects of the planning of the project.

For instance: how should we get our sample of students so that this purpose could be met? And what kinds of tests should the battery include? And how can the resultant data be used? Let's consider these three questions in turn, starting with the problem of sample selection.

SELECTING THE SAMPLE

If we wanted to be sure of having a reasonable number of cases requiring specific talent, we would need a very large group to start with.
And to get accurate census data so that we would know how the various kinds of talents are distributed, in other words what the manpower resources of the United States are, the sample would have to be not only large but also representative. We started with the premise that we could not test more than about one twentieth of the high school students in the United States. Practical considerations imposed this limit. As for the question of how the sample was to be chosen, a sample composed of volunteers—whether volunteer students or volunteer schools—was clearly out, and so was a sample based on the quota method. The limitations of these procedures are well known. They result in samples that are entirely or largely self-selected on the basis of convenience and availability, and are thus not at all likely to be representative. And we needed a sample that we could have confidence was adequately representative, not only in regard to a few specified characteristics for which quotas might be established, but in regard to all characteristics. This requirement dictated the need for some sort of randomization procedure, preferably stratified random sampling.

Having decided this, the next problem was the choice of the kind of sampling unit to be used. The primary sampling unit might have been the student, the school, a specific grade in a specific school, the school system, the county, or even the state. If the first of these alternatives, the individual student, were the sampling unit, each student would be chosen at random individually to be in the sample.

The fact that any particular student had been chosen would have no bearing on whether any of his classmates or schoolmates, neighbors or people in his school or in the state, or any fellow-residents of his town, county, or state were also chosen. At the other extreme, if the state were the primary sampling
unit, it would mean that perhaps one or two or more states would be
chosen at random from among the fifty, and that then only students in
the selected states would be tested. This would not have been a
good way of doing the sampling for Project Talent, since there are
everseous differences among states in regard to educational standards
and many other highly relevant factors. A small handful of states
could hardly be expected to represent the country as a whole very well.

Using the individual student as the sampling unit would have made
it possible to draw accurate conclusions about American high school
students in general on the basis of a smaller sample than any other
procedure would have required. But from any realistic viewpoint it
would not have been feasible to test individual students on a large
scale. For this and other reasons it was decided to use whole schools
as the sampling unit. More specifically, in the case of the public
schools, in situations where junior high schools existed, the sampling
unit was to be the senior high school together with its associated
junior high school(s). (One purpose of treating junior high schools
this way was to maximize the extent to which all four grades (9 through
12) would be included for the same school system.)

The purpose of using stratified random sampling rather than
simple random sampling was to increase the efficiency of the sample,
by making it as representative as possible in regard to certain important
variables. The stratification variables were to be variables that
there might be some reason to expect would be related to level of
test scores of the students in the school. This relation might exist
either because the chosen stratification variable was somehow related
to the average ability level of the pupils or because it was related
in some important way to the nature of the school program.
The initial stratification of the schools was on the basis of whether they were public, parochial, or private. Public schools were then stratified on the basis of three variables: school size, location, and extent of drop-outs.

School size was considered important because, as Conant has pointed out, if a school is very small it cannot offer as varied a curriculum as a larger school, except at much greater cost per capita. Accordingly, the public high schools were divided into four strata in terms of Grade 12 enrollment (or Grade 11 in the case of those few school systems in the South which still have an eleven-year program):

1) Grade 12 enrollment: under 25
2) Grade 12 enrollment: 25-99
3) Grade 12 enrollment: 100-399
4) Grade 12 enrollment: 400 and above

Another advantage of this split on school size is that it would help insure appropriate representation of urban and rural schools, since most of the extremely small schools would be in rural areas.

The record stratification variable was location. Broad geographical region was an important consideration because of regional differences in the economy, the nature of the population, and the character of the schools. The actual state in which the school was located was also important, because of the functions of most state departments of education in setting standards for the schools and the teachers, requirements for high school diplomas, etc. Another factor to be considered in connection with location was the distinction between very large cities and other types of local units, including smaller cities and villages. The geographical stratification procedure that
was followed was one that had the advantage of taking into account these three important considerations--broad geographical region, specific state, and the distinction between very large cities and other local units. What was done was to divide the country into 56 geographical units: the five cities with populations in excess of 1,500,000 (New York, Chicago, Los Angeles, Philadelphia, Detroit); the 50 states excluding the five cities mentioned above; and the District of Columbia. These 56 geographical units were then grouped into nine broad geographical regions.

The third stratification variable for public schools was extent of school drop-outs. The proportion of students that drop out of high school before graduating is obviously a significant factor, a high drop-out rate being almost universally regarded as undesirable. A "retention index," representing the tendency for the students to remain in school until they graduate, was therefore computed for each high school, to serve as a basis for stratification. The retention ratio was defined as the ratio of number of graduates to number of tenth-graders.

The parochial schools and the private schools were stratified into the 56 geographical units, but unlike the public schools, they were not stratified on school size or retention ratio, because there were too few of them.

The next basic question was what "sampling ratio" should be used, in other words what proportion of the schools should be in the sample. It was decided to aim at testing somewhere between 400,000 and 500,000 students. There were several reasons why so large a number seemed desirable. In the first place the decision to use the school as the basic sampling unit necessitated a larger sample than would have been
required if the individual student had been the sampling unit. Furthermore, the initial sample would have to be extremely large because it would eventually be broken into so many different kinds of subdivisions that some of these smaller groups would be very small indeed.

In view of the fact that some of the occupations that are extremely important from the standpoint of the welfare of the nation (for instance nuclear physicist) would ultimately attract only a very small percentage of the high school students tested, a very large number of cases would be necessary in order to obtain enough cases in these important but relatively infrequent occupations that worthwhile information would be yielded about them. And even in the case of more usual professions, such as dentist, where the total number of persons would undoubtedly be larger, they might have to be divided into many sub-groups, homogeneous in terms of type of community, family background, high school curriculum, size of high school, sex of the student, and other factors. These sub-groups might be too small to provide useful information unless the initial size of the Project Talent group was very large indeed. This was perhaps the major consideration that led to the decision to aim at testing close to half a million students.

The biggest remaining problem was the question of how to get an adequate representation of the larger public high schools. There are a great many more very small high schools in the United States than very large ones. If a uniform sampling ratio were used for all kinds of high schools, and if this sampling ratio were set to give the desired total number of high schools or the desired total number of students, the sample would be very heavily loaded with extremely small schools and would have far too few of the very large schools to make it possible to determine the effects of school size on test results or on subsequent
activities of the student, or, for that matter, on anything else that might be of concern. The solution lay in the use of differential sampling ratios. It was decided, therefore, to use a basic sampling ratio of one in 20 for medium-sized public schools, a much larger sampling ratio, one in 13, for the largest public schools, and a much smaller sampling ratio, one in 50, for the smaller ones.

Naturally it was planned that the schools would be weighted appropriately in the analysis of the data, to adjust for the differential sampling ratios, since only in this way can results be obtained which permit sound inferences about the total population of high schools or high school students. The sample was actually selected by a computer, the IBM-650, which generated a series of random numbers for the purpose.

THE TESTS

So much for the sampling. The next problem was the nature of the tests. Since the purposes of Project Talent were broad, a wide variety of tests would be needed. Accordingly the final battery included 18 tests, yielding over 40 scores.

INFORMATION TEST

One of the tests which accounted for about 30 per cent of the testing time and has yielded at least 16 scores per student, is the Information Test. (These 16 scores are only a start. Many more scores will be obtained later from this test.)

Because so many scores are derived from the Information Test, it is especially important to understand the role of this test in the battery. The test consists almost wholly of very brief five-choice items testing factual information (including vocabulary). One of its functions is to indicate the amount of information the student has in many specific areas. In several of these areas, amount of information
possessed has been found to be a good index of aptitude. In the case of
the non-academic areas, the information may have been acquired through
participation in the activity as a hobby, extracurricular activity, house-
hold chore, or part-time job, or through independent reading, or special
training, or in some other way. People tend to devote time and effort
to activities in areas in which they are especially interested. Thus
there is a tendency to acquire more information about such areas than
about areas in which one has no particular interest. This principle can
be utilized in interpreting the subscores on the Information Test,
especially in the case of subjects in which the high school offers no
formal instruction, and subjects in which the courses offered are largely
electives. For this purpose, it is necessary to take into account not
only the individual scores, but also the complete pattern of scores, and
also to bear in mind that differential interests are not solely responsi-
ble for different amounts of information in different areas since not
everyone has good opportunity to acquire information in a particular area.

Certain subscores on the Information Test are useful as direct
indications of the amount of information the student has acquired in
certain areas of an academic nature; for instance, physical science,
biological science, social studies, literature. In the academic areas,
while many of the test items are concerned with information that is
normally covered in the school curriculum, some students will be able to
answer many questions on the basis of information acquired outside of
school (for instance, through independent reading, listening to the radio,
watching television).

The Information Test has 395 items, and yields scores for informa-
tion in the following areas:

   Literature
   Music
   Social Studies
In addition, it yields a vocabulary score, which indicates the relative size of a student's general vocabulary, a score on a screening scale, a scientific attitude score, and a total score. I'll come back to the scientific attitude score in a few minutes.

The screening scale consists of about a dozen extremely easy items, such as "How many days are there in a week?" interspersed among the regular items. The assumption is that anyone who can read knows the answers to these very easy questions. The purpose of the scale is to identify students who are functionally illiterate, foreign students whose command of English is inadequate for them to do themselves justice on the tests, etc.

The total score, which represents breadth of general information, is the sum of the 15 other scores that have been mentioned. Total score on an information test may be regarded as one very good indicator of general intelligence level, at least in the case of students whose environmental background has been normal. However students from extremely deprived backgrounds, who have not had average opportunities to acquire a broad range of information, will probably have lower total scores than they otherwise would.

I'd like to say a word or two now about some of the Information Test scales that are probably of particular concern to this group -- the Mathematics and science scales for instance. The items in the Mathematics scale are concerned with definitions, the
vocabulary of mathematics, mathematical notation, other kinds of factual information; and the understanding of mathematical concepts. None of the items requires computation or reasoning or calls for solving a problem, since these capabilities are covered by other tests in the battery--the Arithmetic Computation Test and the Mathematics Test.

The Physical Science scale includes items about chemistry, physics, astronomy, and other physical sciences.

The Biological Science scale includes items about botany, zoology, and microbiology. A few items about nature lore are included, though most of the items are concerned with more formal aspects of biological science.

The Aeronautics and Space items are on such topics as piloting procedures, navigation, jet planes, and space exploration. Much of the information that the student has in this area is likely to have been acquired out of school.

The items on electricity and electronics stress information that is acquirable through direct experience in the construction and maintenance of electrical and electronic equipment. Students who have worked on radios, hi-fi sets, or other electronic equipment, or on mechanisms with electric motors should get good scores. A good score on this scale is likely to indicate both interest and aptitude.

The mechanics items tap a wide range of information, mostly concerned with common machines and tools with which boys who are interested in mechanical activities are likely to be familiar. The emphasis is on information that is likely to be acquired through direct experience with tools, engines, and motors. It is reasonable
to assume that anyone who has both aptitude and interest in this area would have acquired considerable information of the type tested, even if he attends a high school that does not offer formal instruction in mechanical subjects.

The Scientific Attitude scale, which I mentioned earlier, is a brand-new type of scale, and it is in the battery for strictly experimental purposes. It is hoped that the score on this scale will be indicative of what a student's general attitude is, in the face of unexplained phenomena. For instance does he view the world as a place where there are logical cause-and-effect relationships or, on the other hand, does he regard it as a place where consequences are illogical, capricious, and arbitrary. Students who are superstitious should get low scores on this scale. Among these students will be those who believe such things as that black cats bring bad luck, horseshoes bring good luck, Friday the 13th means trouble, there is something in astrology (at least when the astrologer is an expert), those who believe that the world is full of evil spirits that have to be propitiated, etc. Those who favor anthropomorphic explanations of the behavior of lower animals should also tend to get low scores, as should those who do not seek a logical explanation for unexplained phenomena, but instead tend to accept naively the first explanation presented that is superficially attractive.

Among the people who we hope will get high scores are those who, like the scientist, seek reasonable and logical explanations for events and unexplained phenomena, and tend to reject explanations that involve occult arts and magic. It should be noted that a person may have a scientific attitude even if he has never studied science formally and has little or no factual information about how scientific research is actually carried out. It will be seen that the trait to be differentiated by these scores
involves not only the judgment and reasoning power necessary to decide whether a "conclusion" follows logically from given premises but also a basic tendency to use the scientific approach.

Each item consists of a description of a phenomenon or an occurrence. Five explanations are presented, and the task is to select the "best" one. Only one of the five explanations is reasonable and logical. Of the other four, some involve superstitions, and concepts incompatible with the scientific viewpoint. Selection of other distractors, however, may primarily imply muddled thinking on the part of the examinee. None of the items requires possession of any special information in science or mathematics. Whenever such information is needed to answer the question, it is presented in the item itself. Here is a sample item.

Professor Rogers wished to find out whether any of the 950 students in Central High School could demonstrate the power of "mind over matter." When ten pennies are tossed, the chances that all ten of them will fall "heads up" are about one in a thousand. Rogers had each student in turn toss ten pennies. He instructed them to try, by thinking very hard about it, to make all ten pennies fall "heads up." But when one of the boys, Joe Thompson, tossed the coins they all fell "tails up." What does this suggest about Joe?

A. Joe was purposely trying to get all tails.
B. Joe became confused.
C. Joe didn't have faith in the power of mind over matter.
D. Joe is unlucky.
E. Nothing.

(The last one is the answer.)

It is hypothesized that scores on this test are related to science aptitude.

**ACHIEVEMENT AND APTITUDE TESTS**

In addition to the Information Test, the achievement-and-aptitude part of the battery includes an achievement test in English, an achievement test in mathematics other than arithmetic computation, a separate
Arithmetic Computation Test, and 14 other tests most of which should be regarded primarily as measures of aptitudes.

ACHIEVEMENT TESTS

The English test yields five separate subscores (spelling, capitalization, punctuation, English usage, and effective expression) and a total score.

The Mathematics test is in three parts. Part I is the Arithmetic Reasoning section, in which a systematic effort was made to eliminate almost all computation, even at an easy level. This was done in an effort to reduce the correlation between scores on this test and scores on the Arithmetic Computation test—a correlation which is usually quite substantial. The aim of the Arithmetic Reasoning items is solely to determine whether the student can do the reasoning part of a problem—not whether he will do the subsequent computation carefully enough to avoid careless errors, nor whether he knows that \(9 \times 7 = 63\). Measuring these things is the function of the Arithmetic Computation test. Several devices were used to avoid contaminating the Arithmetic Reasoning test with arithmetic computation skills. One such device was to require the student to tell how he would solve a stated problem but not to require him to actually do the necessary computation. Here is a sample item:

Three percent of a certain number is 141. To find the number, you would

A. divide 141 by .03
B. divide 141 by 3
C. divide 141 by 300
D. multiply 141 by .03
E. multiply 141 by 300
Here is another sample:

Mrs. Rogers buys $\frac{6}{10}$ pounds of apples.  
She gives the clerk a one-dollar bill.  
What single other fact is needed to find  
out how much change she should get?

A. Price per apple  
B. No. of apples per pound  
C. No. of ounces in a pound  
D. Price per pound of apples  
E. Weight of average apple

No items were included which required algebraic techniques for  
solution. However, no problem solvable without algebra was proscribed  
just because it could be solved efficiently by algebra, provided there  
was an alternate way of solving it, without algebra.

It may appear to some that the type of item used requires substantially  
more verbal ability than the more conventional type of arithmetic reasoning  
item. But there is no good reason for believing this is so. Verbal ability  
is necessarily involved in comprehending any verbally expressed arithmetic  
reasoning problems, whether or not computational ability is also involved.

Eliminating the computational requirement does not increase the absolute  
amount of verbal ability required. It merely increases relative importance  
of verbal ability in comparison with computational ability.

Part II of the Mathematics test measures achievement in all kinds of  
mathematics generally taught up to and including the 9th grade, with the  
exception of the areas covered in the Arithmetic Computation Test and in  
Mathematics Part I (Arithmetic Reasoning). The primary emphasis of this  
test is on elementary algebra; other topics include fractions, percents,  
square roots, intuitive geometry, and elementary measurement formulas.

The sum of the Part I and Part II scores gives an indication of achieve-  
ment in mathematics through the Grade 9 level. It should be indicative of  
aptitude for further work in mathematics.
Part III of the Mathematics test covers topics normally taught in grades 10-12 in college-preparatory courses. As in the case of Mathematics Part II, and the items are intended primarily to test understanding and application of basic concepts and methods, not rote memory. Areas sampled include plane geometry, solid geometry, algebra, trigonometry, elements of analytic geometry, and introductory calculus. While some of these subjects are not offered in most high schools, students who have taken college-preparatory mathematics beyond the 9th grade level and have really understood it should be able to score well.

A senior who gets an extremely high score on the first two parts of the mathematics test and can't do much with Part III is likely to turn out to be someone who dropped math after the ninth grade even though he had excellent aptitude for it.

A total Mathematics score, representing the sum of all three parts, is also obtained.

**APTITUDE TESTS**

Among the aptitude tests are two tests which measure certain kinds of memorization ability. One is the Memory for Sentences test and the other is the Memory for Words test. The latter is designed primarily to measure one aspect of aptitude for learning foreign languages. Other aspects of foreign language aptitude are measured by the Disguised Words test and the Word Functions in Sentences test.

The remaining tests in the battery are:

1. Reading Comprehension
2. Mechanical Reasoning
3. Two tests of spatial visualization ability (visualization in Two Dimensions and Visualization in Three Dimensions)
4. Abstract Reasoning
   This is a non-verbal test to measure one kind of abstract reasoning ability, involving diagramatic materials.
5. Three tests of clerical and perceptual speed and accuracy.

(These are the Table Reading Test, the Clerical Checking Test, and the Object Inspection Test.)

6. The Preferences Test.

(This is another test that was included solely for experimental use.)

7. Creativity Test.

(This is a test that probably has considerable relevance to science aptitude.)

Each item in the Creativity test consists of a complex problem similar to a practical problem that might be encountered in real life. The examinee must think of a clever or ingenious solution. The five choices are given in terms of the first and last letters of possible right answers. This is to insure that the examinee really develops the solution, instead of just picking it from among five alternatives.

Here is a sample item.

The sanding and smoothing of knobs and other small round wooden objects is a problem because flat sheets of sandpaper do not fit the knobs. The sandpaper buckles and tends to wrinkle. One solution is to cut several

A. e - - - - - n d - - s in the sandpaper.
B. h - - - - - l t - - s in the sandpaper.
C. p - - - - - l s - - s in the sandpaper.
D. t - - - - - n w - - s in the sandpaper.
E. m - - - - - h b - - s in the sandpaper.

(The answer is C, since "parallel slits" is the solution.)

That is the battery. Now I'd like to say a word or two about a couple of actual cases, to give some idea of how the test results can be used.

USE OF RESULTS

Let's start with Larry, a very bright eleventh-grader almost 17 years old, who doesn't seem to have any definite career plans. Larry's pattern
of scores suggests considerable ability in math and science. His Information Test scores in Math and Physical Science place him in the top one per cent of eleventh graders. And he answered nine of the ten scientific attitude items correctly. His scores on the Mathematics achievement test are also extremely high—even his Part III score which includes some Grade 12 work. Although he has had only \(2\frac{1}{2}\) years of high school math he got 9 out of 14 items right on Part III, which puts him at the 97th percentile for eleventh-graders and at the 94th percentile in comparison with twelfth-graders. He seems not to be much interested in things mechanical. His Mechanics information score is only at the 29th percentile for boys. His Electricity and Electronics score is at the 38th percentile. His father is an engineer but Larry's interests don't seem to lie along these lines. He says he expects to become a dentist but that he would really like to go into a different field—something not included in a fairly comprehensive list showing 35 categories of occupations. In his theme on "My views about an ideal occupation" he says he wants to do graduate work—field unspecified—and also be an amateur magician. Well, maybe there is no necessity for Larry to have his career plans clearly mapped when he is only in the eleventh grade. But he does seem to have excellent aptitude for science and math. Let's hope that if he isn't aware of this—and he may not be—he knows it by the time he definitely has to choose the career he will train for.

The second student I want to tell you about is Andy. Andy is a senior. He is at the 99th percentile on Math I and II, but he appears not to have had much math beyond the ninth grade. If he had had three or four years of high school math, surely with his mathematical ability he would have been able to answer more than seven of the 14 items correctly in Part III. Eventually we will have norms on Part III of the Math test for students who have had specified amounts of high school math. These norms will be useful in cases such as Andy's. Meanwhile his scores point up the need for early identification of talents.