

STANDARDIZED TESTS

W. C. HAWTHORNE, CRANE JUNIOR COLLEGE, CHICAGO

Tests or examinations of some sort are a recognized part of the instruction in every serious course of study. If it is a standard course,—one in which the same topics, essentially, are given to large numbers of classes, the tests in the different classes will necessarily be somewhat alike. If a test covers materials commonly given in all schools, and if we *know*, not merely guess, what the average student should do on such a test,—better still, if we know what the best ten percent will do, the next best ten percent, and so on, we shall have a Standardized Test.

The preparation of a Standardized Test involves much more work than is given commonly to a set of ordinary examination questions. The subject matter of the test should cover not only what has just been studied but what we may reasonably expect will be studied next year and every year in all schools. This will make a long examination, you may say. It is true, but not a long one for the pupil to answer, or for the teacher to score, as I shall show. Moreover, the Standardized Test parts company with the idea that any pupil shall answer all the questions. If the test is to be used in many different schools, in order to give every one a chance, it should contain from twenty-five to fifty questions, correct answers for sixty percent of which may, perhaps, be regarded as a high score.

Such a test as this is scored easily if the answers are very short. There are several types of questions admitting of such answers, and they need not be of the "yes or no" type either. First may be placed numerical problems, and the pupil should have been trained to indicate the process by which the answer was obtained; not the literal formula, but one containing all the numerical data, with the operations indicated. Perhaps the pupil is called upon to indicate, by underscoring, which of several answers given is the correct one. Perhaps certain statements are given, to be checked as true, false or uncertain. I think the "true-false" test, and especially the

usually recommended method of summing up the scores (i. e. rights minus wrongs) is about discredited. Not only may guessing play too large a part in the student's performance, but the subtraction of the number of "wrongs" from "rights" may leave him with a minus score. Moreover, it can be proved that the subtraction of all or any part of the wrong answers from the number of right ones leaves the examinee with exactly the same relative rank in the class as when the usual method is followed,—counting up the correct answers only. But if a third type of statement be added, the truth or falsity of which it is impossible to determine from the data given, and if, moreover, both false and uncertain statements are so worded as to have a considerable air of plausibility, the possibility of getting a high score by guessing is probably no greater in this than in any other test. Still another type of easily scored question is that in which the one word which gives the correct meaning to a sentence is omitted. Considerable skill is needed here, if the missing word is not to be too obvious on the one hand, or the meaning of the sentence, on the other hand, too obscure even to those well acquainted with the subject. It is obvious that the wording of questions must be such that teachers cannot disagree on the scoring of any question.

The arrangement of the questions will depend upon two points of view; whether we wish to test the student's quickness to recall and use facts which should be thoroly known; or to determine his grasp of the subject,—his power. If the former, we will make the questions in any set uniformly difficult, and impose a sharp time limit, so that only the quickest will be able to finish; if the second, we will let the questions in each set increase sharply in difficulty, so that only those who have the greatest mental power will be able to answer all. Plenty of time should be allowed on this test, and in order that those who get through first shall not disturb the others, additional questions should be provided, the scores of which, however, should not be recorded with the others. At Crane Technical High School and Junior College, we have now in preparation physics tests in sets of twenty-

five questions, with as many alternative sets as possible, of which the high school pupils are expected to answer the first twenty, and the college students the whole number. Not infrequently the self complacency of some careless collegian has received a severe jolt, when we have shown him that his score on a certain test is less than that of the average high school boy.

We are now ready to give the test for the first time. It is evident that the usual precautions to prevent copying and communication are more necessary when the answers are as short as in this sort of a test. There should be a printed or mimeographed sheet of the questions for each student. Distribute them face downward, not to be looked at until the signal to begin is given. I have not been able to discover that it makes any difference whether the answers are written on the same or another sheet of paper. If the questions are supposed to be of about equal difficulty, note the time of starting, and stop when the first few have finished. This gives an idea of the standard time to be fixed for the test.

Before scoring, prepare a key on a slip of stiff cardboard, with answers so placed that they correspond in position with the answers on the students' papers. If the test is answered by underlining words, or filling in blanks, the key may be made on a sheet of celluloid or transparent paper, or a cardboard with holes at the proper places to show the correct answers. With such devices as these, the drudgery of scoring the papers is reduced to a minimum, and it is very evident that more questions per tests, and more tests per year, can be handled without adding to the already heavy burden of the conscientious teacher. Add to this the moral satisfaction of knowing that you have eliminated from your grade book the influence of the famous "sweet smile and winning way", and you will be willing to take the extra trouble necessary in preparing the tests.

Although in assigning values to correct answers it is very easy to say that students should be marked as right or wrong only, in practice it is often very difficult to do. Suppose a pupil has worked a problem by the right method, but because of a mistake in arithmetic, or a mis-

take in copying, he puts down the wrong answer. Should he not get some credit for it? Many questions, again, can not be answered properly by the use of several words. Suppose the full answer is "Simple harmonic motion" and the student writes "Harmonic motion." It is neither completely right nor completely wrong. It has therefore been my custom to give *two* for correct answers, *one* those partly right. Elaborate methods have been worked out for giving weighted scores, the values of which are proportional to the percentage of pupils failing on each question. It is doubtful if such a scheme is worth the trouble it takes. Better recast the whole test if there is any great difference in the difficulty of the questions.

Having scored the papers, we must next study them for certain necessary information. To find whether the test is too hard or too easy, we plot a curve, using the scores as abscissas, and the number of pupils making each score as ordinates. An approximation to the normal probability curve shows that we have a test reasonably suited to the capacity of our class; if it is skewed to the right, it is too easy; if to the left, it is too hard.

We may next study the relative difficulty of the questions by recording the number of times each question has been missed. If any question has been missed by no one, it is too easy and should be dropped. On the contrary, if any question is missed by all, it is too hard; the form is ambiguous, or the content is too difficult for their comprehension, or the subject has not been taught properly. It should be dropped or reshaped. If the number of correct answers to each question runs ninety-five to five percent, we may conclude that we have a good set.

The next step is to revise the questions, eliminating some, inserting others, restating some, and putting the harder questions toward the end. Another trial on the same or a similar class should be much more satisfactory; that is, it should give a nearly normal distribution in which the failures should be much more numerous in the latter part of the test than in the first part.

We are now ready to compare the grades on this particular set of questions with a criterion. About the only

one we have is the teacher's judgment, as derived from his knowledge of their laboratory work, their written reports, their daily recitations and the average of all their previous tests to date. If in general the best students have made the best scores and vice-versa,—in other words if there is a high correlation, say 0.6 or more, between the class grades and the results of this test, we may conclude that we have a satisfactory test. Standardization may then proceed as fast as we can give the test to other classes in the same or other schools, accumulate and classify the scores, and establish norms.

What about the student's reaction to this type of tests? Uniformly satisfactory. They do not dread them; they are even eager for them. It is amusing to see them "get set" and "go" after a few experiences, and settle down to a good natured race to get every word possible written in the ten or fifteen minutes before "time" is called. They feel that the larger number of questions gives them a better chance of telling what they know, and there is never any chance for a difference of opinion as to what a certain answer is worth.

Finally, no claim is made that all the testing of a year's work can be done with this kind of tests. Even for the measurable results of our work, an occasional one or two hour test of the old type is desirable. And no scheme of "testing", no matter how perfected, should blind our eyes to the most important objectives of our work:—the appreciation and love of science, the habit of mental honesty, and of deferring judgment until all available evidence is weighed carefully; in a word, that attitude of mind that will make of the raw material that comes to us good citizens and reverent, obedient inhabitants of a Universe that is governed by Law. These are things that can not be measured by an examination nor expressed by grades in a class-book.