SCORING THE TEST

Scoring Objective Test Items

If the student's answers are recorded on the test paper itself, a scoring key can be made by marking the correct answers on a blank copy of the test. Scoring then is simply a matter of comparing the columns of the answers on this master copy with the columns of answers on each student's paper. A strip key which consists merely of strips of paper, on which the columns of answers are recorded, may also be used if more convenient. These can easily be prepared by cutting the columns of answers from the master copy of the test and mounting them on strips of cardboard cut from manila folders.

When separate answer sheets are used, a scoring stencil is more convenient. This is a blank answer sheet with holes punched where the correct answers should appear. The stencil is laid over the answer sheet, and the number of the answer checks appearing through holes is counted. When this type of scoring procedure is used, each test paper should also be scanned to make certain that only one answer was marked for each item. Any item containing more than one answer should be eliminated from the scoring.

As each test paper is scored, mark each item that is scored incorrectly. With multiple choice items, a good practice is to draw a red line through the correct answers of the missed items rather than through the student's wrong answers. This will indicate to the students those items missed and at the same time will indicate the correct answers. Time will be saved and confusion avoided during discussion of the test. Marking the correct answers of the missed items is simple with a scoring stencil. When no answer check appears through a hole in the stencil, a red line is drawn across the hole.

In scoring objective tests, each correct answer is usually counted as one point, because an arbitrary weighting of items make little difference in the students' final scores. If some items are counted two points, some one point, and some half point, the scoring will be more complicated without any accompanying benefits. Scores based on such weightings will be similar to the simpler procedure of counting each item on one point. When a test consists of a combination of objective items and a few, more time-consuming, essay questions, however, more than a single point is needed to distinguish several levels of response and to reflect disproportionate time devoted to each of the essay questions.

When students are told to answer every item on the test, a student's score is simply the number of items answered correctly. There is no need to consider wrong answers or to correct for guessing. When all students answer every item on the test, the rank of the students' scores will be same whether the number is right or a correction for guessing is used.

A simplified form of item analysis is all that is necessary or warranted for classroom tests because most classroom groups consist of 20 to 40 students, an especially useful procedure to

compare the responses of the ten lowest-scoring students. As we shall see later, keeping the upper and lower groups and ten students each simplifies the interpretation of the results. It also is a reasonable number for analysis in groups of 20 to 40 students. For example, with a small classroom group, like that of 20 students, it is best to use the upper and lower halves to obtain dependable data, whereas with a larger group, like that of 40 students, use of upper and lower 25 percent is quite satisfactory. For more refined analysis, the upper and lower 27 percent is often recommended, and most statistical guides are based on that percentage.

To illustrate the method of item analysis, suppose we have just finished scoring 32 test papers for a sixth-grade science unit on weather. Our item analysis might then proceed as follows:

1. ... Rank the 32 test papers in order from the highest to the lowest score.

2. ... Select the 10 papers within the highest total scores and the ten papers with the lowest total scores.

3. ... Put aside the middle 12 papers as they will not be used in the analysis.

4. ... For each test item, tabulate the number of students in the upper and lower groups who selected each alternative. This tabulation can be made directly on the test paper or on the test item card.

5. ... Compute the difficulty of each item (percentage of the students who got the item right).

6. ... Compute the discriminating power of each item (difference between the number of students in the upper and lower groups who got the item right).

7. ... Evaluate the effectiveness of distracters in each item (attractiveness of the incorrect alternatives).

Although item analysis by inspection will reveal the general effectiveness of a test item and is satisfactory for most classroom purposes, it is sometimes useful to obtain a more precise estimate of item difficulty and discriminating power. This can be done by applying relatively simple formulas to the item-analysis data

Scoring Essay Type Test Items

According to N.E. Gronlund (1990) the chief weakness of the essay test is the difficulty of scoring. The objectivity of scoring the essay questions may be improved by following a few rules developed by test experts.

a. Prepare a scoring key in advance. The scoring key should include the major points of the acceptable answer, the feature of the answer to be evaluated, and the weights assigned to each. To illustrate, suppose the question is "Describe the main elements of teaching." Suppose also that this question carries 20 marks. We can prepare a scoring key for the question as follows.

i. Outline of the acceptable answer. There are four elements in teaching these are: the definition of instructional objectives, the identification of the entering behaviour of students, the provision of the learning experiences, and the assessment of the students' performance.

ii. Main features of the answer and the weights assigned to each.

- Content: Allow 4 points to each elements of teaching.

 - Comprehensiveness: Allow 2 points.

- Logical organization: Allow 2 points.

- Irrelevant material: Deduct upto a maximum of 2 points.

- Misspelling of technical terms: Deduct 1/2 point for each mistake upto a maximum of 2 points.
- Major grammatical mistakes: Deduct 1 point for each mistake upto a maximum of 2 points

- Poor handwriting, misspelling of non-technical terms and minor grammatical errors: ignore. Preparing the scoring key in advance is useful since it provides a uniform standard for evaluation.

b. Use an appropriate scoring method.

There are two scoring methods commonly used by the classroom teacher. The point method and the rating method. In the point method, the teacher compares each answer with the acceptable answer and assigns a given number of points in terms of how will each answer approximates the acceptable answer. This method is suitable in a restricted response type of question since in this type each feature of the answer can be identified and given proper point values. For example: Suppose that the question is: "List five hypotheses that might explain why nations go to wars." In the question, we can easily assign a number of point values to each hypothesis and evaluate each answer accordingly. In the rating method, the teacher reads each answer and places it in one of the several categories according to quality. For example, the teacher may set up five categories: Excellent – 10 points, good – 8 points, average – 6 points, weak – 4 points and poor – 2 points. This method is suitable in an extended response type of question since in this type we make gross judgment concerning the main features of the answer. It's a good practice to grade each feature separately and then add the point values to get the total score.

- Read a sampling of the papers to get a 'feel' of the quality of the answers. This will give you confidence in scoring and stability in your judgment.
- Score one question through all of the papers before going on to the next question. This procedure has three main advantages. First, the comparison of answer makes the scoring more exact and just, second, having to keep only one list of points in mind saves time and promotes accuracy and third, it avoids halo effect. A halo effect is defined as the

tendency in rating a person to let one of its characteristics influence rating on other characteristics.

- Adopt a definite policy regarding factors which may not be relevant to learning outcomes being measured. The grading of answer to essay questions is influenced by a large number of factors. These factors include handwriting, spelling, punctuation, sentence structure, style, padding of irrelevant material, and neatness. The teacher should specify which factor would or would not be taken into account and what score values would be assigned to or deducted from each factor.

- Score the papers anonymously. Have the student record his name on the back or at the end of the paper, rather than at the top of each page. Another way is to let each student have a code number and write it on his paper instead of his name. Keeping the author of the paper unknown will decrease the bias with which the paper is graded