

answering each item, especially in case of essay type questions. So that the test maker should carefully judge the amount of time taking the types of items, age and ability of the students and the nature of the learning outcomes expected. Experts are of the opinion that it is better to allow more time than to deprive a slower student to answer the question.

Instructions about basis for answering

Test maker should provide specific direction on the basis of which the students will answer the item. Direction must clearly state whether the student will select the answer or supply the answer. In matching items what is the basis of matching the premises and responses (states with capital or country with production) should be given. Special directions are necessary for interpretive items. In the essay type items clear direction must be given about the types of responses expected from the pupils.

Instruction about recording answer

Students should be instructed where and how to record their answers. Answers may be recorded on the separate answer sheets or on the test paper itself. If they have to answer in the test paper itself then they must be directed, whether to write the correct answer or to indicate the correct answer from among the alternatives. In case of separate answer sheets used to answer the test direction may be given either in the test paper or in the answer sheet.

Instruction about guessing

Direction must be provided to the students whether they should guess on uncertain items or not in case of recognition type of test items. If nothing is stated about guessing, then the bold students will guess these items and others will answer only those items of which they are confident. So that the bold pupils by chance will answer some items correctly and secure a higher score. Therefore a direction must be given 'to guess but not wild guesses.'

Preparing the Scoring Key : A scoring key increases the reliability of a test. So that the test maker should provide the procedure for scoring the answer scripts. Directions must be given whether the scoring will be made by a scoring key (when the answer is recorded on the test paper) or by a scoring stencil (when answer is recorded on separate answer sheet) and how marks will be awarded on the test items.

In case of essay type items it should be indicated whether to score with the 'point method' or with the 'rating method.' In the 'point method' each answer is compared with a set of ideal answers in the scoring key. Then a given number of points are assigned. In the rating method the answers are rated on the basis of degrees of quality and determine the credit assigned to each answer. The scoring key helps to obtain a consistent data about the pupils' performance.

the test maker should prepare a comprehensive scoring procedure along with the test items.

TRY OUT OF THE TEST

Once the test is prepared now it is time to be confirm the validity, reliability and usability of the test. Try out helps us to identify defective and ambiguous items, to determine the difficulty level of the test and to determine the discriminating power of the items. Try out involves two important functions.

(a) *Administration of the test.*

(b) *Scoring the test.*

(a) **Administration of the test.** Administration means administering the prepared test on a sample of pupils. So the effectiveness of the final form test depends upon a fair administration.

Gronlund and Linn have stated that 'the guiding principle in administering any class room test is that all pupils must be given a fair chance to demonstrate their achievement of learning outcomes being measured.' It implies that the pupils must be provided congenial physical and psychological environment during the time of testing. Any other factor that may affect the testing procedure should be controlled.

Physical environment means proper sitting arrangement, proper light and ventilation and adequate space for invigilation. Psychological environment refers to those aspect which influence the mental condition of the pupil. Therefore steps should be taken to reduce the anxiety of the students. The test should not be administered just before or after a great occasion like annual sports on annual drama etc. One should follow the following principles during the test administration.

- The teacher should talk as less as possible.
- The teacher should not interrupt the students at the time of testing.
- The teacher should not give any hints to any student who has asked about any item.
- The teacher should provide proper invigilation in order to prevent the students from cheating.

(b) **Scoring the test :** Once the test is administered and the answer scripts are obtained the next step is to score the answer scripts.

A scoring key may be provided for scoring when the answer is on the test paper itself. Scoring key is a sample answer script on which the correct answers are recorded.

When answer is on a separate answer sheet at that time a scoring stencil may

be used for answering the items. Scoring stencil is a sample answer sheet where the correct alternatives have been punched. By putting the scoring stencil on the pupils answer script correct answer can be marked.

For essay type items separate instructions for scoring each learning objective may be provided. It has been discussed in detail in the previous section 'Preparation of the scoring key'.

Correction for guessing: When the pupils do not have sufficient time to answer the test or the students are not ready to take the test at that time they guess the correct answer, in recognition type items. In that case to eliminate the effect of guessing the following formula is used

$$\text{Score} = R - \frac{W}{n-1}$$

Where

R = No. of right responses

W = No. of wrong responses

n = No. of alternatives.

But there is a lack of agreement among psychometricians about the value of the correction formula so far as validity and reliability are concerned. In the words of Ebel "neither the instruction nor penalties will remedy the problem of guessing."

Guilford is of the opinion that "when the middle is excluded in item analysis the question of whether to correct or not correct the total scores become rather academic." Little said "correction may either under or over correct the pupil's score."

Keeping in view the above opinions, the test-maker should decide not to use the correction for guessing. To avoid this situation he should give enough time for answering the test item.

EVALUATING THE TEST

Evaluating the test is most important step in the test construction process. Evaluation is necessary to determine the quality of the test and the quality of the responses. Quality of the test implies that how good and dependable the test is? (validity and reliability). Quality of the responses means which items are misfit in the test. It also enables us to evaluate the usability of the test in general class-room situation. Evaluating the test involves following functions.

- (a) Item analysis.
- (b) Determining validity of the test.
- (c) Determining reliability of the test.
- (d) Determining usability of the test.

(a) Item analysis

Item analysis is a procedure which helps us to find out the answers to the following questions.

- = Whether the items functions as intended ?
- = Whether the test items have appropriate difficulty level ?
- = Whether the item is free from irrelevant clues and other defects ?
- = Whether the distractors in multiple choice type items are effective ?

The item analysis data also helps us

- to provide a basis for efficient class discussion of the test result
- to provide a basis for the remedial works
- to increase skill in test construction
- to improve class-room discussion.

Item Analysis Procedure : Item analysis procedure gives special emphasis on item difficulty level and item discriminating power. The item analysis procedure follows the following steps.

1. The test papers should be ranked from highest to lowest.
2. Select 27% test papers from highest and 27% from lowest end. For example if the test is administered on 60 students then select 16 test papers from highest end and 16 test papers from lowest end.
3. Keep aside the other test papers as they are not required in the item analysis.
4. Tabulate the number of pupils in the upper and lower group who selected each alternative for each test item. This can be done on the back of the test paper or a separate test item card may be used (Fig. 3.1)
5. Calculate item difficulty for each item by using formula

$$\text{Item difficulty} = \frac{R}{T} \times 100$$

Where R = Total number of students got the item correct.
T = Total number of students tried the item.

In our example (fig. 3.1) out of 32 students from both the groups 20 students have answered the item correctly and 30 students have tried the item. The item difficulty is as following :

$$\begin{aligned} \text{Item difficulty} &= \frac{R}{T} \times 100 \\ &= \frac{20}{30} \times 100 = 66.67. \end{aligned}$$

It implies that the item has a proper difficulty level. Because it is customary to follow 25% to 75% rule to consider the item difficulty. It means if an item has a difficulty more than 75% then it is a too easy item if it is less than 25% then it is a too difficult item.

6. Calculate item discriminating power by using the following formula.

$$\text{Item discriminating power} = \frac{R_U - R_L}{T/2}$$

Where R_U = Students from upper group who got the answer correct.

R_L = Students from lower group who got the answer correct.

$T/2$ = half of the total number of pupils included in the item analysis.

In our example (Fig. 3.1) 15 students from upper group responded the item correctly and 5 from lower group responded the item correctly.

$$\begin{aligned} \text{Discriminating power} &= \frac{R_U - R_L}{T/2} = \\ &= \frac{15 - 6}{16} = \frac{9}{16} = .56 \end{aligned}$$

A high positive ratio indicates the high discriminating power. Here .56 indicates an average discriminating power. If all the 16 students from lower group and 16 students from upper group answer the item correctly then the discriminating power will be 0.00. It indicates that the item has no discriminating power. If all the 16 students from upper group answer the item correctly and all the students from lower group answer the item incorrectly then the item discriminating power will be 1.00 it indicates an item with maximum positive discriminating power.

7. Find out the effectiveness of the distractors. A distractor is considered to be a good distractor when it attracts more pupils from the lower group than the upper group. The distractors which are not selected at all or very rarely selected should be revised. In our example (fig. 3.1) distractor 'D' attracts more pupils from.

Fig. 3.1 The item analysis card.

Subject		Unit							
Class									
Item No.									
Group	No. of Pupils	Alternatives					Total No. of pupils responded	Item difficulty	Item Dis. power
		A	B	C*	D	E			
Upper	16	0	0	15	1	0	16	66.67	.63
Lower	16	5	4	5	0	0	14		

*Correct answer.

upper group than the lower group. It indicates that distractor 'D' is not an effective distractor. 'E' is a distractor which is not responded by any one. Therefore it also need revision. Distractor 'A' and 'B' prove to be effective as it attracts more pupils from lower group.

Preparing a test item file : Once the item analysis process is over we can get a list of effective items. Now the task is to make a file of the effective items. It can be done with item analysis cards. The items should be arranged according to the order of difficulty. While filing the items the objectives and the content area that it measures must be kept in mind. This helps in the future use of the item.

(b) **Determining Validity of the test :** At the time of evaluation it is estimated that to what extent the test measures what the test maker intends to measure. We shall discuss in detail different methods of determining validity of a test in chapter—5.

(c) **Determining Reliability of the test :** Evaluation process also estimates to what extent a test is consistent from one measurement to other. *Otherwise the* results of the test can not be dependable. The methods of determining the reliability of a test is discussed in detail in chapter—5.

(d) **Determining the usability of the test :** Try out and the evaluation process indicates to what extent a test is usable in general class-room condition. It implies that how far a test is usable from administration, scoring, time and economy point of view. For detail description about usability see chapter—5.

MODEL QUESTIONS

1. What is a standardized achievement test ? Write briefly the procedure of preparing a standardized achievement test.

2. Discuss the steps you follow in constructing a test.
3. How do you prepare a good test ?
4. What are the general principles of test construction ?
5. What is item analysis ? Discuss the process of item analysis in detail.
6. Discuss in detail the need for try out of a test before its standardization.

□□□