Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology

Suruchi
Research scholar, M. D. University Rohtak, Haryana.

Surender Singh Rana
Associate Professor& Principal, T. R. college of Education, Sonepat, Haryana.

Current piece of work is focussed on the analysis of test items constructed in the subject of Biology for class IX. It involved the analysis of difficulty level and discrimination power of each test item. There were 120 objective type questions in the achievement test and was administered on a sample of 500 students from government and private schools. The findings of the work showed that most of the test items fall in the acceptable range of difficulty index and discrimination index. However, nine of the test items were rejected due to very high or very low difficulty level and poor discrimination power. Using the findings relationship between the difficulty index of each item and the corresponding discrimination index is carried out using Pearson correlation formula. Instead of a linear relation, it was found that item discrimination power increased with the increase in difficulty value but got decreased for very easy and very difficult test items. This work is significant for the researchers and school teachers in framing test items with optimum difficulty and discrimination power.

INTRODUCTION
Achievement tests are one of the most important aspect of teaching – learning process and the two most important characteristics of an achievement test are its reliability and content validity. For a test to be reliable and valid, a systematic selection of items with regard to subject content and degree of difficulty is necessary. Moreover, the reliability of the test also depends upon the grading consistency and discrimination between the students of different performance levels. Thus the quality and effectiveness of a test depends upon the individual item. To determine the quality of individual item, item analysis is done after the administration and test depends upon the individual item. To determine the quality of individual item, item analysis is done after the administration and scoring of the preliminary draft of the test on the selected sample. Ebel1972, “Item analysis indicates the difficulty level of each item and discriminate between the better and poorer examinees. According to Brown and Frederick1971, item analysis has two purposes: First, to identify defective test items and secondly, to indicate the content the learners have or have not mastered. Item analysis measures the effectiveness of individual test item in terms of its difficulty level and power to distinguish between high and low scorers in test. Thus it helps in selecting and retaining the best test items in the final draft of the test rejecting poor items and also show the need to review and modify the items.

OBJECTIVES
1. To find out the item difficulty level and discrimination power of individual test item.
2. To find out the relationship between degree of item difficulty and corresponding power of discrimination of test items.

SAMPLE
Random sampling method was adopted to select a sample of 500 secondary school students. Sample included the students of both genders.

DATA COLLECTION
Researcher herself conducted the test on a sample of 500 students from government and private schools. Students were given as much time as they required to complete the test. Students were instructed to fill their answers only in the answer sheet provided.

STATISTICAL ANALYSIS
After scoring the test items, test scores were arranged in descending order using MS-EXCEL. To carry out the item analysis top 27% scorers and bottom 27% scorers of the total examinees were selected. Upper and lower 27% was used as “this value maximize the difference in normal distributions while providing enough cases for analysis” (Wiersma& Jurs, 1990, p145). Middle 46% of the test scores were excluded as they behave in a similar pattern contributing insignificantly to discriminate the performance by students. Nunnally(1972) suggested 25% while SPSS (1999) uses the highest and lowest one third (33%).

Difficulty value (p): Acc. to Frank S. Freeman2 “Difficulty value of an item may be defined as the proportion of certain sample of subjects who actually know the answer of an item.” Index of difficulty for each test item can be calculated as

\[ D_v = \frac{R_v}{N_u + N_l} \]

\[ D_v = \frac{\sum (R_u + R_l)}{N_u + N_l} \]

Dv = item difficulty
Rv = the number of students in the upper 27% who responded correctly
Higher the values of the difficulty index, easier the item is. Brown (1983) – In item difficulty, if most students answered an item correctly then the item was an easy one. If most students answered an item incorrectly then it should have been a difficult one. So, the items answered correctly by 100% or 0% of the examinees are insignificant.

Discrimination power (d)
Blood and Budd (1972) defined the index of indiscrimination as the ability of an item on the basis of which the discrimination is made between superiors and inferiors.

\[
D_p = \frac{R_u - R_l}{0.5 N}
\]

\(D_p\) – discrimination power
\(N\) – total no. of correct responses
\(R_u\) = the number of students in the upper 27% who responded correctly
\(R_l\) = the number of students in the lower 27% who responded correctly

Numerical value of discrimination power may range from -1.00 to +1.00. In 1986, Ebel and Frisbie gave following rule of thumb for determining the quality of items with respect to their discrimination index.

If a test item is correctly answered by upper group students and incorrectly by lower group students than the item is said to have positive discrimination power. If a test item is correctly answered by lower group students and incorrectly by upper group students than the item is a negative discriminator. Item with negative discrimination decreases the validity of test and thus must be discarded. If a test item is answered correctly by equal number of upper and lower group students than its showing zero discrimination.

Relationship between item difficulty (p) and discrimination power index (d) for each test item was determined by Pearson correlation analysis. The difficulty indices and discrimination indices are most often reciprocally related.

FINDINGS AND DISCUSSION
The findings of item analysis on 120 test items can be understood by developing histograms for difficulty index and discrimination index of test items.
Histogram-2 showing discrimination power of test items clearly shows that 81% of the test items had a discrimination index of .40 and above and fall in the category of excellent items. Fourteen items are considered as good with a Dp range of .30 to .39 while two items with Dp range between .20 to .29 needed improvement and six items which fall below .20 of discrimination index were out rightly rejected. None of the test item showed negative discrimination. One item was found to have zero discrimination power.

Scatter diagram showing the correlation between Dv and Dp is given below.

CONCLUSION
Findings of this study signifies the importance of item analysis for determining the quality and utility of individual test item in constructing a more reliable test. The study suggests that test items with good positive discrimination power and moderate difficulty are ideal for a good test whereas the items with negative or zero discrimination power and having very low or very high difficulty level should be out rightly rejected. Item analysis results are tentative and are influenced by the number and kind of students, instructional procedure applied, chance errors and purpose of the test. As, in a class test very difficult or very easy test items may be retained despite of their poor discrimination power because the purpose is to test the content mastery and attainment of set objectives and not to discriminate the superior and inferior students. To develop a perfect test is almost an undoable task but test analysis provides an empirical data about the quality of items that can be quite significant in improving the evaluation process. Heretofore item analysis is quite significant in developing a worthy test.

As the scatter dia. is showing, the relationship between “p” and “d” is not linear but is somewhat dome shaped. Pearson correlation coefficient “r” calculated as -0.3711 showing a moderate negative relationship between values of Dv and Dp. This negative correlation signifies that as the difficulty index increased discrimination index also increase but to an optimum value only after which discrimination power decrease with the increase in difficulty level. This suggested that the easier items ( >.80 ) or too difficult items ( <.20) poorly discriminate between the superior and inferior examinees. Mitra et al; (2009) reported poor correlation of discrimination index with difficulty index. Sim and Rasiah (2006) found the maximum discrimination power of items with the difficulty index between .40 to .74. In the present work also maximum discrimination was observed in the difficulty index range of 0.5 to 0.7. Items with similar difficulty, if show, scattered discrimination values then it shows the guessing practices by the examinees. However in this study no wide scattering was observed. Thus, after the item analysis of test items nine items were recommended to reject for the final drafting of achievement test.