Research Paper Education

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.

ABSTRACT

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.

KEYWORDS Item Analysis, Difficulty level, Discrimination index.

INTRODUCTION

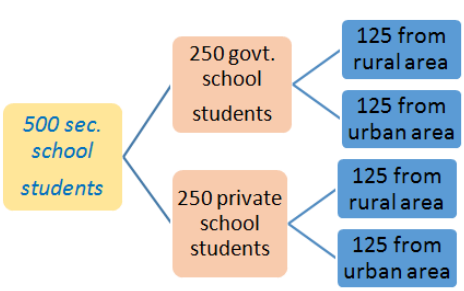
Achievement tests are one of the most important aspect of teaching – learning process and the two most important characteristics of an achievement test are its reliability and content validity. For a test to be reliable and valid, a systematic selection of items with regard to subject content and degree 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 scoring of the preliminary draft of the test on the selected sample. Ebel[1] 1972, "Item analysis indicates the difficulty level of each item and discriminate between the better and poorer examinees. According to Brown and Frederick[2], 1971, 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.



INSTRUMENT USED

An Achievement test of 120 objective test items was used for data collection. The test is developed by the researcher with the help of subject experts on the selected three chapters in Biology subject for class IX as per CBSE course. Blueprint of the test was prepared giving due weightage to the instructional objectives, subject content and forms of questions. Bloom's revised taxonomy was used to frame the test items. Test comprised of eighty multiple choice questions, fifteen true-false type questions, eighteen sentence completion type and seven matching type questions. While constructing the items it was ensured that language of test items was simple and unambiguous providing no clues towards the correct answer. Instructions were made clear on the test and a separate answer sheet was used to fill the answers.

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[3], 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[4](1972) suggested 25% while SPSS (1999) uses the highest and lowest one third (33%).

Difficulty value (p): Acc. to Frank S. Freeman[5] "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

$$Dv=(R_u + R_l) / (N_u + N_l)$$

Dv = item difficulty
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
N_u = the number of students in the upper group
N_l = the number of students in the lower group

Difficulty Value	Quality	Recommendation
• below 0.20	• vey difficult/ misleading	• Discard
• 0.20 to 0.50	• Good	• Retain
• 0.50 to 0.80	• Best	• Retain
• above 0.80	• very easy/ poor item	• Discard

Higher the values of the difficulty index, easier the item is. Brown[7] (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[8] (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.

$Dp=(R_u - R_l)/ 0.5 N$
Dp – 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[9] gave following rule of thumb for determining the quality of items with respect to their discrimination index.

Discrimination Index	Quality of item	Recommendation
above 0.40	Excellent	Retain
0.30 to 0.39	Good	scope of improvement
0.20 to 0.29	Mediocre	Need to review
below 0.20	poor	Discard
Negative value	worst	Definitely discard

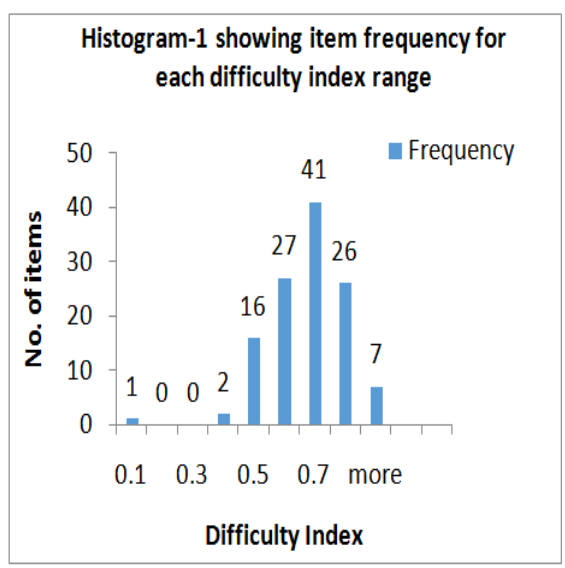
If a test item is correctly answered by upper group students and incorrectly by lower group students than the item is said to have positive discrimination power. If a test item is correctly answered by lower group students and incorrectly by upper group students than the item is a negative discriminator. 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 oftenreciprocally related.

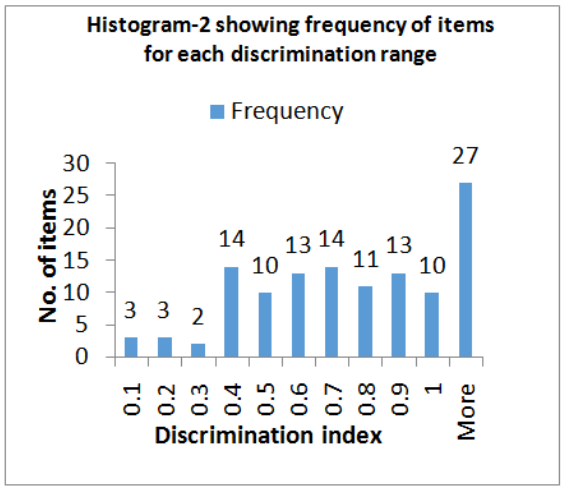
Pearson correlation cofficient (r)	Kind of relationship
+ 0.70 or higher	very strong positive relationship
+.40 to +.69	strong positive relationship
+.30 to.39	moderate positive relationship
+.20 to .29	weak positive relationship
+.01 to .19	No or negligible relationship
-.01 to -.19	No or negligible relationship
-.20 to -.29	weak negative relationship
-.30 to -.39	moderate negative relationship
-.40 to -.69	strong negative relationship
-.70 or higher	very strong negative relationship

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.

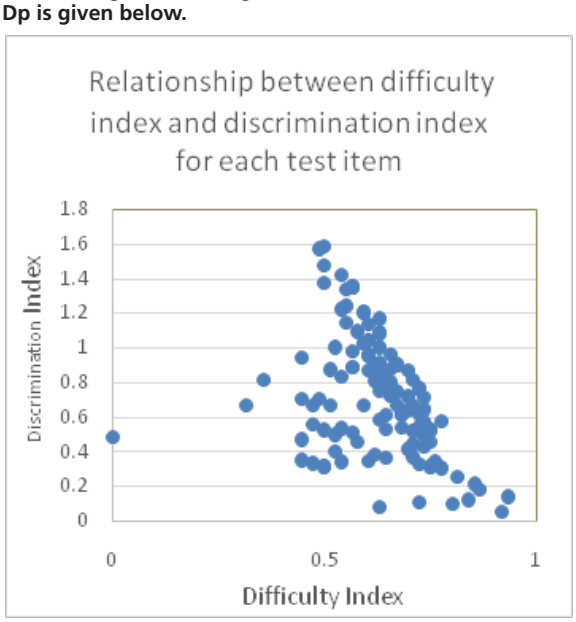


In histogram-1, showing difficulty indices for 120 items, only one item falls below .20 while eighteen items falls in the range of 0.20 to 0.50 of difficulty index. A total of 94 items (78%) are in the range of 0.51 to 0.80 and seven items are found to be very easy with difficulty value of above 0.80. One very difficult and seven very easy test items were rejected for the final draft of achievement test.



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.



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[9]etal; (2009) reported poor correlation of discrimination index with difficulty index. Sim and Rasiah[10],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 to0.7. 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.

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.

REFERENCES

1. Ebel, R.L. (1972) Essentials of Educational Measurement (1st Edition). New Jersey: Prentice Hall. | 2. Brown, Frederick G., "Measuring Classroom Achievement", Halt Richard and Winston, U.S.A.", c1981, pp. 101-110, 224p. | 3. Wiersma, W. &Jurs, S. G. (1990). Educational measurement and testing (2nd Ed.). Boston, MA: Allyn and Bacon. | 4. Nunnally, JC. 1972. Educational measurement and evaluation. 2nd edition. New York. McGraw-Hill. | | 5. Freeman Frank S. (1962).Theory and Practice of Psychological Testing, New Delhi: Oxford &Ibh publishing | | 6. Thompson, B. and Levitov, J. E. (1985). Using microcomputers to score and evaluate test items. Collegiate Microcomputer, 3, 163-168. | 7. Brown, F. (1983), "Principles of Educational and Psychological Testing," Third edition, NY: Holt, Rinehart & Winston. Chapter 11. | 8. Blood, D.F. and W.C. Budd, (1972). Educational measurement and evaluation. New York: Harper and Row. | 9.Ebel, R.L. &Frisbie, D.A. (1991).Essentials of EducationalMeasurement (5th Ed.). New Delhi: Prentice Hall of India Pvt. Ltd. | 10.Mitra,N.K., Nagaraja, H.S., Ponnudurai, G. & Judson, J. P. (2009) The levels of difficulty and discrimination indices in type A multiple choice questions of Pre-clinical Semester 1 multidisciplinary summative tests. IeJSME, 3, 1, pp. 2-7. | | 11.Si-Mui-Sim and Rasiah R.I. (2006) Relationship between item difficulty and discrimination | indices in true/false type multiple choice questionsof a para-clinical multidisciplinarypaper. Ann Acad Med Singapore, 35, pp. 67-71. | |