CONSTRUCTION AND STANDARDIZATION OF ACHIEVEMENT TEST IN MATHEMATICS

Rachhpal Singh

GHG Khalsa College of Education, Gurusar Sadhar, India.

ABSTRACT

In the present study an attempt was made to construct and standardize Achievement test in mathematics for the selected topics from 9th class P.S.E.B. Text-Book. The scale initially consists of 78 Items. After review and evaluation of statements by experts, items were reduced to 70. After item analysis, 40 items were retained in the final draft of achievement test. The reliability and validity of the test was found to be 0.79 and 0.78. Each right answer has 1 mark and wrong answer has 0 marks.

Keyword: Achievement

Introduction

Mathematics is considered as one of the most important subjects at school level all over the world. Teaching mathematics to those students who do not have a strong base in mathematics is challenging. Testing the achievement in mathematics is again is challenging issue as there is lack of such standard tools.Standardized tests make extensive use of problems and assess the child's ability to reason and evaluate situations. If a student has an experiential understanding of math processes, he or she will be able to understand the test questions and apply the proper skill to successfully accomplish the required task. The development procedure of a criterion referenced achievement test to measure achievement in different taxonomic categories differs from norm referenced achievement test.Stanley (1964) has formulated four rules which should be followed

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

139 | P a g e

in planning a test which includes, Adequate provision should be made for evaluating all important outcomes of instructions, The test should reflect the appropriate proportion of emphasis of the course, The nature of the test must reflect its purpose, The nature of the test must reflect the conditions under which it will be administered.These rules were kept in view while constructing the achievement test. The construction of the test was carried out by a consideration of limitations under which the test was to be developed. The investigator had limited time at his disposal. The investigator followed a fixed path and the detailed set of specifications as to the purpose of test and the time, cost and resources at the disposal were taken into account. The length of the test, type of nature of the test-items, and method of scoring the test were also some basic considerations which were planned in advance.

Objective

To construct and standardize Achievement test in Mathematics

Preliminary draft

The construction of the test was carried out with adequate planning. A blue print was prepared from the contents of 9[th]class PSEBtext book revised in 2014. Bloom's taxonomic objectives formed the basis for the blue print of the test. While making the preliminary draft of the test, the existing tests in the concerned Subject were consulted. Many new and original test items were also prepared to assess the desirable objectives of the blue print. A large number of test items of suitable difficulty were compiled and only 78 items were prepared initially. Rough idea of the difficulty of the items was obtained by trying out the items on a small group of students from the population. It was tried out on a sample of 30 students of class IX from GHG Khalsa Senior Secondary School, Gurusar Sadhar district Ludhiana, affiliated to Punjab School Education Board. These students were asked to point out any difficulty in understanding the items. Whenever the students pointed out a difficulty, it was noted down and necessary modifications were made. The items were then edited and carefully worded instructions, which indicated briefly the nature and purpose of the test. The manuscript of statements was given to 08 experts in order to improve the items; comments of the experts were obtained along with their suggestions and opinion. On the basis of their feedback some items were modified and some were deleted. After this exercise second draft of test was finalized, it consists of total 70 items. After the necessary modifications in the light of the experts' suggestions and 'small-group try

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

140 | P a g e

out', the preliminary draft was printed and cyclostyled. For recording the responses of the subjects a separate response sheet was enclosed with the booklet.

Final draft

The selected test items were included in the final draft of the test which was administered to another 100 ninth class students of one school affiliated with P.S.E.B.Rapport was established with the group. They were told the objectives of administering of the test and assured that their responses would be kept confidential and used only for research purpose. Once they were clear about all these, each student was given a copy of test along with a blank answer sheet on which they had to respond. Before responding to the test, the students read the instructions given in the test. They were told that there was no time limit for responding. Each correct answer was awarded 01 mark and incorrect answer 0 marks. Item analysis was done and two kinds of information, namely, item difficulty and discriminating power of items were computed.

Item analysis

For this test,item analysis was carried out in accordance with Kelley's (1939)method. Kelley demonstrated that when extreme groups, each consisting of 27% of the total group were used, the ratio of the difference in abilities of the group to the standard error of their difference, that is, the degree of uncertainty about the size of the real difference, was found to be maximum. Kelley (1939) showed that by taking upper and lower groups of 27% of total sample, one could say with the greatest confidence that those in the upper group were superior in ability measured by the test to those in the lower group.

For calculating the Discriminating Power (D.P.) the following formula, was used

$$DP = \frac{RU-RL}{0.5\ N}$$

For calculating the discriminating value (D.V.) the following Formula, Was used

$$DV = \frac{RU+RL}{N}$$

For the selection of item, the criteria recommended by Ebel (1966)were given due consideration.

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

141 | P a g e

The evaluative criteria for selection of items according to index of discrimination power is given below:

Table 1

Evaluative criteria for selection of test items according to index of discrimination power

Index of Discrimination power	Item evaluation
0.40 and above	Very good item
0.20 to 0.39	Good but marginal usually subject to modification
Below 0.19	Poor item

Table 2

Distribution of discriminating power on the first draft of achievement test in mathematics

Sr. No.	D.P.	F	Test item	Remarks
1	More than or Equal to 0.40	39	3,4,5,7,8,9,10,11,19,20,21,22,24,29,30,35,36,37,39,40,42, 43,46,47,49,50,51,54,55,57,58,60,61,62,63,64,65,66,67.	Test item Accepted
2	0.20 to 0.39	19	6,12,13,15,16,18,23,25,26,27,28,41,44,53,56,59,68,69,70	Subject to modified
3	0.19 & below	12	1,2,14,17,31,32,33,34,38,45,48,52.	Test items rejected

Above table reveals that 39 items have discriminating power equal to 0.40 or above hence these were selected to be included in the final draft, 19 items have D.P. range 0.20 to 0.39. These items were revised and modified and included in the final draft. The remaining 12 items have D.P. below 0.19, hence these were rejected.Test items having D.P. 0.20 to 0.39 were modified in the light of clarity of language, complexity of content and difficulty level of items.

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

142 | P a g e

Table 3

Evaluative criteria for selection of test items according to index of difficulty value

Index of difficulty value	Item evaluation
Above 0.67	Poor item
0.60 to 0.67	Good but marginal usually subject to modification
0.20 to 0.59	Very good item
Below 0.20	Poor item

Table 4

Distribution of difficulty value power of the first draft of problem solving ability test in Mathematics

Sr. No.	D.V.	F	Test Items	Remarks
1.	More than 0.67	16	4,5,6,7,8,9,11,12,13,18,20,23,29,37,69,70.	Rejected
2.	1.60 to 0.67	18	3,14,16,17,25,26,27,28,43,45,51,52,53,54,58,60,66,67.	Subjected to modified
3.	0.20 to 0.59	35	1,2,10,15,19,21,22,24,30,31,32,33,35,36,38,39,40,41,42,44,46,47,48,49,50,54,56,57,59,61,62,63,64,68.	Accepted
4.	Below 0.20	1	34.	Rejected

Thus, the items having difficultyvalue above 0.67 and below 0.20 were rejected.Items having D.V. between 0.20 and 0.59 were accepted as such. The items whose difficulty value was between 0.60 to 0.67 were accepted after revision and modification. Out of 70 items 35 were selected as such, 18 revised and modified, and 17 test items were rejected.In total, on the basis of DV's and DP's test items 30 were rejected and 40 were selected for final draft of the test.

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

143 | P a g e

Reliability of Test

Reliability refers to the consistency of scores or measurement which is reflected in the reproducibility of the scores. A test is said to be consistent over a given period of time when all the examinees retain their same relative ranks of two separate testing with the same test in other words reliability tests that to what extent individual differences of scores can be assigned to chance error. There are various methods to compute the estimates of reliability, but this does not entail that the reliability coefficients obtained with different measures are inter-changeable. The split half method is often regarded as the best of methods for determining test reliability. Perhaps its main advantage is that all the data for determining test reliability are obtained upon one occasion, hence variations introduced by differences between two testing situations are eliminated.

The splitting up of the whole test into two equivalents can be performed in various ways and thus the unique value of reliability coefficient is subject to negation. The commonly is done by splitting the odd and even number of items accurately. Then the correlation between scores on odd and even number items is calculated. Reliability, thus obtained is of half length test. To obtain estimate of reliability of total test it is necessary to correct or step up half the test correlation to the expected full length value. This is done with help of Spearman-Brown prophecy formula. The Spearman - Brown prophecy formula assumes that variances of the two halves are equal. The reliability for the half test ($r_{1/2}$) was found to be 0.66 and the reliability of the whole test thus, computed was found to be 0.79

Validity of test

Validity means truthfulness. The validity of test or any measuring instrument depends upon the validity with which it measures whatever it purports to measure. "A test is valid when the performances which it measures correspond to the same performances as otherwise indecently measured or objectively" (Garrett, 1985).The degree of validity of the single most important aspect of test validity can be best defined as the degree to which a test is capable of achieving certain aims. The validity of the test is determined by measuring the extent to which it matches with a given criterion. The process of validity involves checking the agreement between the responses elicited by each question item against the criterion.The correlation of test scores and marks in mathematics in final standard examination were

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

144 | P a g e

computed. The computed validity of the test was found to be 0.78

References

Best, J. W. &Kahan, J.V. (1995).*Research in Education*., Prentice Hall of India Pvt. Ltd., New Delhi.

Ebel, R.L. (1966). *Measuring educational achievement*, Prentice Hall of India Pvt. Ltd., New Delhi.

Garrett, H.E. (1985). *Statistics in Psychology and Education,*VakilsFeffer and Simons Pvt. Ltd., Mumbai.

Kelley, T. l. (1939). Statistical methods in educational measurement, *Journal of Statistics and operational research*, 9, 17-24

Stanley (1964)."Behavioral Study of Obedience."Journal of Abnormal Social Psychology 67,1-78.

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.
International Research Journal of Human Resources and Social Sciences (IRJHRSS)

145 | P a g e