ISSN: 0975-833X

*Available online at http://www.journalcra.com*

*International Journal of Current Research*
*Vol. 10, Issue, 11, pp.75451-75455, November, 2018*
**DOI: https://doi.org/10.24941/ijcr.33168.11.2018**

**INTERNATIONAL JOURNAL OF CURRENT RESEARCH**

## RESEARCH ARTICLE

# DEVELOPMENT AND STANDARDIZATION PROCESS OF MATHEMATICS ACHIEVEMENT TEST FOR THE STUDENTS OF GRADE X

***Rajendra Kunwar**

Reader, Mathematics Education, Tribhuvan University, Faculty of Education, Mahendra Ratna Multiple Campus, Ilam, Nepal

**ARTICLE INFO**

***Article History:***
Received 10th August, 2018
Received in revised form
27th September, 2018
Accepted 19th October, 2018
Published online 30th November, 2018

***Key Words:***
Achievement test,
Test Development,
Item Analysis,
Test Blueprint.

**ABSTRACT**

This article is concerned with the achievement test on mathematics for grade X students. The test consists of the 50 multiple choice items extracted from the grade X mathematics course. Moreover, the purpose of this study is to develop and standardize the achievement test to measure the students' cognitive level of grade X mathematics. The preliminary instrument consisted of 60 multiple-choice items and first it was tested on a sample size of 50 students of grade X. After the refinement of the test items using different procedures, 50final achievement test items were designed. The second test administration was organized after 40 days in the same group of students. This test construction and development consists of the steps as preparation of test blueprint, preliminary draft and answer key, pilot testing, item analysis, preparation of final test, reliability and validity of the test and norms. Discriminative Index and power of difficulty level was calculated. Test reliability coefficient was calculated and found to be 0.86 and also the internal consistency Cronbach's α was found to be 0.85.Likewise, validity was also established through the expert judgment and establishing reliability coefficient. The age norm was also established.



***Citation: Rajendra Kunwar,* 2018.** "Development and standardization process of mathematics achievement test for the students of grade x", *International Journal of Current Research*, 10, (11), 75451-75455.

## INTRODUCTION

An achievement test is designed to measure knowledge, understanding, and skills in a specified subject or group of subjects." To sum up these, it is understood that this test is a successful means of determining to what extent the student has attained a certain degree of progress toward a desired goal (Freeman, 2010). Achievement tests often measure the learner's cognitive trait and designed to measure subject and grade-level specific knowledge. Achievement test is an important tool in school evaluation and has great significance in measuring instructional progress in the different subject area of the learner (Mehrens and Lehmann, 1973). The term 'achievement' means one's learning attainments, accomplishments, proficiencies, etc. which is directly related to the pupil's status of educational improvement. It is a systematic procedure for determining the amount a student has learned through instruction (Gronlund, 1971). It gives an accurate picture of students' knowledge and skills in the related subject area or domain being tested. The accurate data or achievement of the students is very important for planning, implementing and evaluating the curriculum.

****Corresponding author:*** **Rajendra Kunwar,**
Reader, Mathematics Education, Tribhuvan University, Faculty of Education, Mahendra Ratna Multiple Campus, Ilam, Nepal.

Teachers use achievement tests to measure the students' performance in their area of study. Employers use achievement tests to measure the capabilities of potential employees. Professional associations use achievement tests to select the best fitted qualified applicants for the required profession. In any circumstances the achievement test is used to select the competent persons in the required field. It is used to determine what a person knows about a certain subject or to determine an individual's level of skill in a certain area. An achievement test is designed to measure a person's level of skill, accomplishment, or knowledge in a specific area. The achievement tests constitute an important tool in the school evaluation program. It is necessary for the teacher to distinguish the students' capabilities in particular subject area. The pupils differ in the attainments significant among themselves. In the school evaluation program, various forms of achievement tests are used to measures the accomplishments of the pupils. According to Downie (1961), "Any test that measures the attainments or accomplishments of an individual after a period of training or learning is called achievement test". The achievement testof mathematics constructed in this study was a multiple choice item test. The multiple choice items can measure at both the knowledge and understanding level (Gronlund, 1976; Osadebe, 2013). Moreover, the multiple choice items have the advantage of covering the broad

contents of mathematics curriculum and measures different outcomes of the cognitive level.

### Objectives of the Study

- To develop achievement test covering the whole syllabus of grade X mathematics.
- To standardize the achievement test establishing with difficulty value, discrimination index, reliability, validity and norms.

## MATERIALS AND METHODS

The paper mainly aims at developing and standardizing the achievement test items of mathematics of grade X. It has been made on the basis of the survey carried out among the students of grade X from the community schools located in different district of province No. 1. Moreover, in order to develop and standardize the test items different stepwise processes like test blueprint, pilot testing, item analysis, establishing reliability coefficient, test validation, norms, discriminative index etc. were adopted and finally standardized achievement test items of mathematics for the tenth grade were developed.

### Conceptual Framework

This test was constructed on the basis of secondary level objectives and syllabus of grade X mathematics. The test items were prepared from the prescribed text book of grade X mathematics comparing the objectives and the syllabus. The achievement test is designed to measure students' knowledge, skill, comprehension and application in mathematics. The achievement test was prepared through a systematic process. The process of test development was adopted five basis steps namely: test conceptualization, test construction, item scoring and analysis, reliability and validity and test standardization. The achievement test construction was adopted the following steps as shown below:

**Preparation Test of Blueprint:** A blueprint is a detail plan of any action or outline. It provides the researcher a bird's eye view of the entire test and also provides the users with basis instruction on the rationale for the process in creating test blue print. In education area, blueprint provides students an interactive approach for education planning, curriculum expectation and the learning objectives. A test blueprint ensures appropriate items representation of content with weighting. In blueprint, we can see the content area, unit wise topics, item wise weighting of contents by objectives, number of items and distribution of scores (marks) and their mutual relationships. It is the basis for test construction. To determine the weighting of each given different content areas relating to the objectives of the level, the grid prescribed by Curriculum Development Centre (CDC), Nepal, concerned mathematics teachers and educational experts were taken in to consideration. The details of test after second try out (final preparation) are given in the form of blueprint. The test blueprint is given in table 1below:

**Preparation of Preliminary Draft and Answer Key:** The preliminary draft of achievement test was prepared by the researcher after the finalizing the blueprint. The draft was given to the expert of test construction in education and After receiving their opinions, some corrections and items in difficult language were modified to simple language statement and 10 items were eliminated from the draft. In the item scoring for multiple type questions, answer key ensures the objectivity and the uniformity to the examiner. Likewise, it helps the examiner for their tiny confusion which sometimes occurs.

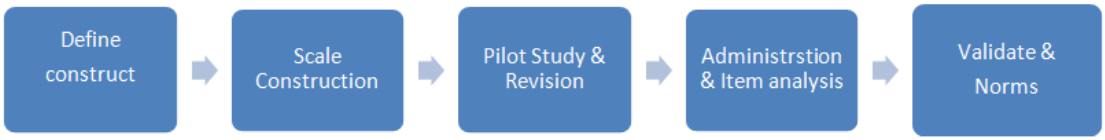


**Figure 1. Steps of Test Construction**

**Table 1. Test Blueprint**

| S. N. | Content Area | Unit | Item wise Weighting of Content by Objectives | | | | No. of Items | Marks |
|---|---|---|---|---|---|---|---|---|
| | | | Knowledge | Comprehension | Application | Skill | | |
| 1. | Sets | 1. Sets | 1 | 1 | - | - | 2 | 2 |
| 2. | Arithmetic | 2. Tax and Money Exchange | 1 | 1 | 1 | - | 3 | 3 |
| | | 3. Compound Interest | 1 | 1 | 1 | - | 3 | 3 |
| | | 4. Population Growth and Development | - | 1 | 1 | - | 2 | 2 |
| 3. | Mensuration | 5. Plane Surface | - | 2 | 1 | - | 3 | 3 |
| | | 6. Cylinder and Sphere | 1 | 1 | - | 1 | 3 | 3 |
| | | 7. Prism and Pyramid | - | - | 2 | 1 | 3 | 3 |
| 4. | Algebra | 8. HCF and LCM | - | 1 | 1 | - | 2 | 2 |
| | | 9. Surds and Radicands | - | 1 | 1 | 1 | 3 | 3 |
| | | 10. Indices | - | 1 | 1 | 1 | 3 | 3 |
| | | 11. Algebraic Functions | 1 | - | 1 | 1 | 3 | 3 |
| | | 12. Equation | 1 | - | 1 | 1 | 3 | 3 |
| 5. | Geometry | 13. Area of Triangle and Parallelogram | 1 | 1 | 1 | 1 | 4 | 4 |
| | | 14. Construction | - | 1 | 1 | - | 2 | 2 |
| | | 15. Circle | 1 | - | 1 | 1 | 3 | 3 |
| 6. | Trigonometry | 16. Trigonometry | 1 | 1 | 1 | - | 3 | 3 |
| 7. | Statistics | 17. Statistics | 1 | 1 | - | 1 | 3 | 3 |
| 8. | Probability | 18. Probability | - | 1 | - | 1 | 2 | 2 |
| | | Total | 10(20%) | 15(30%) | 15(30%) | 10(20%) | 50 | 50 |

**Table 2. Difficulty Level (P-Level) Interpretation Range**

| S. No. | Types of Items | Difficulty Level (P-Level) |
|---|---|---|
| 1. | Long answer type | 50% |
| 2. | Questions with 5 alternatives | 60% |
| 3. | Questions with 4Alternatives | 62% |
| 4. | Questions with 3Alternatives | 66% |
| 5. | Questions with 2 Alternatives | 75% |

**Table 3. Discrimination index (D-Level) Interpretation Range**

| Range of D-Level | Grade | Recommendations |
|---|---|---|
| $D_I$> 0.39 | Excellent | Preserve |
| $D_I$= 0.30-0.39 | Good | Possibilities for improvement |
| $D_I$= 0.20-0.29 | Average | Need to verify/review |
| $D_I$= 0.00-0.20 | Poor | Reject or review in depth |
| $D_I$< -0.01 | Worst | Remove |

**Table 4. Power of Difficulty Level (P-Level)**

| Range of P-Level | No. of Items | Total | Remarks |
|---|---|---|---|
| P≤62% (Accept) | 1,2,3,5,6,7,8,9,10,11,12, 14,15, 16,17,20 21,22,23,24,25,26,27,28,29, 30,31,32,33 34,35,36,37, 38,39.40,41,42,43,44,45, 46 47,48,4950,51,52,53,53,55,56,57,58,59,60 | 58 | 58 items are accepted |
| P≥62% (Reject) | 4,13 (Rejected Items) | 2 | |
| Total Items | | 60 | |

**Table 5. Discrimination index(D-Level)**

| Range of P-Level | No. of Items | Total | Remarks |
|---|---|---|---|
| $D_I$> 0.39 | 1,2,3, 5,6,7,8,9,10, 11,14, 15,16,17 20, 21,22,24,25, 29,35,37,38, 40 44,46,49, 50,51,52,57,58, 59,53 | 34 | 34 items lies in the excellent level |
| $D_I$= 0.30-0.39 | 12,18,19,26,27, 28,30, 31 32,33,36,41,45,47,48,55 | 16 | 16 items lies in the good level |
| $D_I$= 0.20-0.29 | 43,54,56,60 | 4 | 4 items lies in average level |
| $D_I$= 0.00-0.20 | 4, 13,23,34,39,42 | 6 | 6 items lies in poor level |
| $D_I$< -0.01 | | - | |
| Total Items | | 60 | |

**Table 6. Correlation Coefficient Value Description**

| Range of Correlation | Interpretation |
|---|---|
| 0.80 − 1.00 | Very strong positive correlation |
| 0.60 − 0.79 | Strong positive correlation |
| 0.40 − 0.59 | Moderate positive correlation |
| 0.20 − 0.39 | Weak correlation |
| .0.00 − 0.19 | Very weak correlation |

So, in each question 1 mark were given for correct response and 0 for wrong response.

**Pilot Testing:** The pilot testing was done on a sample of 50 grade X students selected on the basis of purposive sampling. The number of the students in the first draft was kept small because it was used to check the ambiguity of items and appropriateness of language. After the administration of preliminary draft, suggestions and observations were taken from these students in relation to understanding of statements, appropriate of language, repetition if any. Some of the statements were modified as per the observation given by the first try out (tested) students. The edited scale constituted sixty items. The pilot test helped the investigator to give the final test.

**Item Analysis:** Item analysis is the item-wise analysis which helps to detect the strength and weakness of each test items. As a rule of thumb, the criterion suggested by Jang and Roussos (2007) that items with means less than 2 and more than 4 are to be rejected and by Jackson (1970) that item standard deviations less than 1 SD < 1) are to be eliminated. Moreover, item-total correlation should not be less than 0.25 Likert (1932) or less than 0.2 or 0.3 were deleted in all the constructs 2005(Field). This stage is most important especially at the time of achievement test development. The two extreme groups, 27 of the high scorer and low scorer formed the criterion group for the calculation of discrimination indices or power of difficulty level of the test items. The following formulae were used to calculate difficulty level.

$$P = \frac{R_U + R_L}{N_U + N_L}$$

$P$=Power of difficulty level
$R_U$= Correct response given by upper group
$R_L$ = Correct response given by lower group
$N_U$= Total number of students in upper group
$N_L$ = Total number of students in lower group

Item analysis is used to calculate difficulty level (P-level) and discriminative value (D-level). Such an exercise helps the test setter to ensure that there is no imbalance in the test paper. During item-wise analysis, the test setter analyses each test item on various parameters stated in the blueprint. In the context of difficulty level, the following difficulty levels are suggested for the selection of questions (Katz, 1959).The process of item analysis is carried out by using two contracting test groups composed from the upper and lower 27% of the examinees or answer sheets on which the items are administered or trial tested. The upper and lower 27% when used are better estimate of the actual discrimination value. They are significantly different and the middle values do not discriminate sufficiently. In other to get the groups, the graded test papers are arranged from the highest score to the lowest score in a descending order or ascending order. The best 27% are picked from the top and the poorest 27% from the bottom while the middle test papers are discarded. In the above table-2, if the multiple choice questions with 4 alternatives has been tested and the difficulty level is calculated at 62 or less, than the question will be selected. But, if the difficulty level is more than 62, than the question will be rejected or removed. Sometime, in multiple choice type questions, little blind guessing is likely to occur. There may be many considered guesses, if every answer given with less than complete certainty is called a guess. So, sometimes students had given answer on guessing. To remove these guessing researcher use the formulas given below:

$$S = R - \frac{W}{N-1}$$

Where,

*S = Correct score*
*R = Number of right response*
*W = Number of wrong responses*
*N= Number of options or choice used in the item*

Ebel and Frisbie (1986) gave the following rule of thumb or range of interpretation of discrimination index for determining the discriminating value of the items, in terms of the discrimination index ($D_I$) or D - Level. Table-3 shows the values $D_I$ and their corresponding interpretation. In the above table-3, if the value of discriminating index $D_I$ is greater than 0.39 than it is considered as best and is between 0.30- 0.39 is considered as possibly needed improvement and is between 0.20-0.29 or below than needs to improve or reject.

**Preparation of Final Test:** After first try out of the mathematics achievement test was given and administered in another group of 50 students of grade X. Same process of first try out was followed for finding difficulty value and discriminating index. Tryout is an essential step in the construction and standardization of a test. It helps to detect the difficulty and discriminating power of the test whereby the test items may be arranged in a sequential manner. The investigator established a rapport with the respondents and explained them about the purpose of the test. Then the instruction and statement of test was explained in the class. The marking scheme was used in two forms: Correct answer was given one marks, wrong answer was given zero mark. The number of question after calculating P- Level and D- Level are given in Table- 4 and Table-5. The items lying in excellent and good level were selected for the final draft.

**Reliability of the Test:** A reliability coefficient is a coefficient of correlation between two sets of test scores. According to Anastasi and Ubrina (1982), reliability refers to the consistency of scores obtained by the same persons when they are re-examined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions. Often this is obtained when a particular group of examinees provides scores on two equivalent tests. If equivalent tests are not available, or cannot be administered conveniently, reliability may be estimated by administering the same test after an interval of time. Alternatively, a test may be split into two equivalent parts and obtain the correlations between scores the parts. The reliability of the test was estimated by test-retest method. The test was administered and repeated on the same group of 50 students after a time interval of 40 days. Coefficient of correlation was calculated between the first and second set of scores. The reliability coefficient between two tests was found to be 0.86. Similarly, Cronbach α was also calculated using the formula $\alpha = 2[1-(\sigma^2 odd + \sigma^2 even)$ $\sigma^2$ total). The scores of all 50 students on 50 items were calculated and Cronbach's α was found to be 0.85. Correlation is an effect size and so we can verbally describe the strength of the correlation using the guide that Evans (1996) suggests for the absolute value of r as given in Table-6. In most situations a good achievement test will have a reliability coefficient of .90 or higher. In this case, the reliability coefficient 0.86 lies on very strong positive correlation range on the above table-5.Likewise, the Cronbach's α = 0.85 indicates very good internal consistency in the test.

**Validity of the Test:** The validity coefficient is sometimes said to show how precisely it measures what it ought to measure. According to Anastasi (2007), the validity of a test concerns with what the test measures and how well it does so. In most achievement test, validity is primarily a matter of content or face validity, and only secondarily, a matter of empirical demonstration. Validity must be built into most achievement tests. The content to be covered by an achievement test and the tasks to be used to indicate achievement are best determined by a detailed judgment of the experts.

The validity coefficient reported here show that the achievement test used in the study is reasonably valid. Content validity of the test, which requires the determination of the adequacy of each item, was ensured through careful planning of the test, satisfying the adequacy of sampling of test items models of the construct to be measured and the meticulous analysis of the test items of experts. The statistical validity of the test items included in the test was compared with the objectives of the topics taught. It was found that there was positive correlation between the items and objectives. So it can be clam that test has got content validity. The validity coefficient reported here show that the achievement test used in the study is reasonably valid.

**Norms:** A standardized test must have norms which should be the average performance of a group or groups that the examiner has taken for the administration of the final form. Norms are measure of achievement which represents the typical performance of a group or groups. Norms are used for interpreting the scores of the individual or a class. Norms are empirically established by determining what parsons in a representative group actually do on a test. This achievement test follows the age norms.

**Conclusion**

Achievement tests have been used for the certification and selection to assess students' progress at the end of secondary school and for admission to higher education. It has been widely used in different occupation and specialized areas as well. An achievement test is also used for guidance and counseling. It has found useful in remedial teaching programs as well as in determining the class to which a student should be admitted into. It can be used for formative and summative evaluation program. Administration of these tests at regular intervals is helpful to teacher in knowing the kinds of difficulties faced by the pupils in learning. The test is mainly constructed to evaluate the students completing the course of grade X in the prescribed secondary level curriculum of Nepal. In the Nepalese context, this sort of object (multiple choice type) questions may be more powerful to evaluate the students achievement level. So finally it can say that the achievement test may use as the evaluation instrument for the students of mathematics at grade X.

**Acknowledgements**
This article took its shape in course of my PhD research. I particularly acknowledge to Prof. Dr. Lekhnath Sharma, my PhD supervisor, for his valued insights, suggestions and feedback for shaping the article as such.

## REFERENCES

Anastasi, A. and Urbina, S. 2007. *Psychological testing* (7th Ed.) (New Delhi: Pearson Education Inc. by (Darling Kindersley (India) Pvt. Ltd.,),

Downie, N. M. 1961. *Fundamentals of measurement*. Oxford University Press, New York

Ebel, R.L. and Frisbie, D.A. 1986. *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.

Evans**,** J.D. 1996. Straightforward statistics for the behavioral sciences. USA: Brooks/Cole Publishing, Pacific Grove.

Field, A. 2005b. *Discovering statistics using* SPSS (2nd ed.). London: Sage.

*Freeman* S. 2010. *Theory and practice of psychological testing*, (Third Edition). USA: Oxford & IBH Publishing Co. Pvt. Ltd.

Gronlund, N.E. 1976. *Measurement and evaluation in teaching.* New York: Macmillan Publishing Co. Inc.

Guilford, J.P. 1971. *Psychometric methods*. New Delhi: Tata McGraw-Hill Publishing Company Ltd.

Jackson, D. N. 1970. A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2). New York: Academic Press.

Jang, E. E. and Roussos, L. A. 2007. An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement.* (Vol. 2).

Joe, A.L. 1995. *Basic concepts of educational measurement and evaluation.* Port Harcourt: Paragraphics.

Katz, D. and Stotland, E. 1959. A preliminary statement to a theory of attitude, structure and change. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York: McGraw Hill.

Kline, R. B. 2010. *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Likert, R. 1932. *A technique for the measurement of attitudes*. Archives of Psychology.

Mehrens, A. and Lehmann, J. 1973. *Measurment and evaluation in education and psychology.* New York: Holt, Rinehart and Winston.

Nunnally, J.C. 1981. *Psycholometrictheory*. New Delhi: Tata McGraw-Hill Publishing Company Ltd.

Obioma, G.O. 1985. *Development and validation of a diagnostic mathematics test for junior secondary schools*. Unpublished Doctoral Dissertation, University of Nigeria, Nsukka.

Osadebe, P.U. 2013. Evaluation techniques, DELSU.*Journal of Educational Research and Development*, *12*(1).

*******