International Journal of Social Science Research and Review

http://ijssrr.com
editor@ijssrr.com
Volume 9, Issue 5
May, 2026
Pages: 113-125

Construction and Standardization of Criterion Referenced Achievement Test for Class-VIII Students

Pulak Bhattacharyya*[1]; Prof. Manashee Gogoi[2]

[1] Ph.D. Research Scholar, Department of Education, Dibrugarh University, India

[2] Professor, Department of Education, Dibrugarh University, India

bhattacharyyapulak294@gmail.com*

http://dx.doi.org/10.47814/ijssrr.v9i5.3336

Abstract

This study emphasizes the development and standardization of a Criterion-Referenced Achievement Test (CRAT) designed for students in Class VIII. The main goal of the test was to evaluate students' understanding of specific learning outcomes aligned with the set curricular objectives, rather than to compare their performance with that of their peers. The process of creating the test included meticulously identifying content areas, defining instructional objectives, and designing a blueprint to guarantee a balanced representation of topics and cognitive levels. A set of test items was created in accordance with the blueprint, prioritizing clarity, relevance, and alignment with the intended learning outcomes. To ensure content validity, subject matter experts and experienced educators reviewed the test items, assessing their adequacy, representation, and appropriateness. Modifications were made based on their feedback to ensure that the test adequately covered the intended content domain. The preliminary version of the test was administered to a sample of Class VIII students under standardized conditions. For reliability estimation, the test-retest method was employed by administering the same test to the same group after a specific time interval. The obtained reliability coefficient was found 0.77 which indicating a satisfactory level of consistency and stability over time. Item analysis was conducted to determine the difficulty level and discrimination power of each item, leading to refinement and selection of the final set of test items. The standardized test thus developed is considered a valid and reliable tool for measuring students' achievement in relation to defined criteria.

Keywords: *Standardization, Criterion Referenced Achievement Test, Validity, Reliability*

Introduction

Test is defined as a series of questions on the basis of which some information is sought. An educational test is a standardized procedure to measure qualitatively or quantitatively one or more than one aspect of a trait by means of a sample of verbal or nonverbal behaviour (Singh, 2006). Kaplan and Saccuzzo (2001) opined that a psychological or educational test is a set of items designed to measure

characteristics of human beings that pertain to behaviour. A test is a standardized procedure for sampling behaviour and describing it with scores or categories (Cullari, 1998). A test is an assessment tool or standard procedure used to measure a sample of behaviour by posing a set of questions (Dickson et al., 2020).

Academic achievement is the progress achieved toward the goal of obtaining educational skills, resources, and information. It usually encompasses a variety of disciplines. It is related to success in academic contexts rather than the general acquisition of knowledge in non-academic settings (Bolt, 2011). In the field of educational psychology, academic achievement refers to a student's level of competency in academic subjects as a whole or in a particular subject, like arithmetic or reading. The outcomes of standardized ability tests and performance evaluations by teachers or other supervisors typically serve as indicators of future academic success (American Psychological Association, 2018). The evaluation of every aspect of learning is done by using verbal or written assessment (Apino & Retnawati, 2017).

An individual's performance in any educational test is recorded in terms of the test scores. The first way to compare an examinee's test score with the score of a specific group of examinees on that test is known as norm referencing. The second way of interpreting a test score is to establish an external standard or criterion and compare the examinees test score with it. This process is known as criterion referencing. In a criterion referenced test, there is a fixed performance criterion (Singh, 2006). Criterion-referenced test may be defined as one in which the test performance is linked or related to some behavioural measures or referents (Glaser, 1963).

Criterion referenced assessment (CRA) is the process of evaluating (and grading) the learning of students against a set of pre-specified qualities or criteria, without reference to the achievement of others (Brown, 2002). Criterion-referenced tests are used to determine what students can do and what they know based on a predetermined, specific set of educational outcomes. These outcomes can be determined by the instructor, school, district, or state based on the curriculum standards that are set. Criterion-referenced tests do not compare students to other students, which is the purpose of norm-referenced tests. As long as the criterion-referenced test is properly aligned to the expected educational outcomes, it can give detailed information about how well a student has performed on each outcome included on the test (Bond, 1995). Criterion-referenced tests are developed by reviewing a set of objectives or a curriculum and then composing the test questions with a goal of having the test "determine how well the students have mastered the identified objectives or curriculum. As with teacher-made tests, a criterion-referenced test can contain words that are unusual or rare in everyday speech and reading, as long as they occur in the curriculum and as long as the students have had an opportunity to learn them. With a criterion-referenced test, we are not much interested in differentiating students by their scores" (Bracey, 2000)

To ensure that the constructed test is of quality, the test must follow a standardized process of evaluation and scoring, so as able to determine a student's knowledge about something (Quaigrain & Arhin, 2017). Therefore, as an assessment tool to obtain data about students' progress, test-quality should be well organized to the stipulated curriculum so that it can be used in improving the current learning system (Pandra et al., 2017). The test construction includes a set of detailed processes (Mamolo, 2021). The test makers should start from constructing the variable to be measured. The test is said to be valid if its construct has successfully been articulated and can already capture the idea (Facione et al., 2000). An Achievement Test not only can assist in grading, tracking, placing, promoting and graduating decisions but are also used in identifying the strengths and weaknesses of a program (Mamolo, 2021).

The preparation and standardization of the Achievement Test consisted of four major phases such as planning, construction, evaluation and validation (Chaudhary & Tyagi, 2017). According to Singh & Sharma (2024) preparation of Achievement Test consist of three steps i.e, item writing, checking by expert and item editing. Singh (2006) has enlisted 7 (seven) general steps of test construction such as planning of the test, writing items of the test, preliminary administration (or the experimental try-out) of

the test, reliability of the final test, validity of the final test, preparation of norms for the final test and preparation of manual.

For appropriate planning of the test, the investigator must keep following aspects in mind such as: to whom, what, when and how to measure. It includes designing the test and preparation of the blue print (Chaudhary & Tyagi, 2017). Test preparation involves steps taken to construct and standardize the test. In test preparation, the developer decides on the test format to choose. Test format refers to the type of questions the test will contain (usually one format per test for ease of test takers and scoring). Test formats have two elements: stimulus (e.g., a question or phrase) and mechanism for response (e.g., multiple choice, true false). Test formats may be objective or subjective (Odinaka et al., 2016).

Multiple choice questions are used widely in schools to assess students. A typical MCQ consists of a question or an incomplete statement, and a set of two or more options that consist of possible answers to the question. The student's task is to select the one option that provides the best answer to the question (Quaigrain and Arhin, 2017). Multiple choice questions (MCQs) are being increasingly used in constructing a test. Item analysis examines student responses to individual test items (MCQs) to assess the quality of those items and test as a whole to improve/revise items and the test (Singh, et al., 2009). A good item can assess cognitive, affective, as well as psychomotor domain (Gajjar, et al., 2014).

Ease of scoring can make multiple-choice (MC) testing particularly appealing to teachers who teach courses with large enrolments. Another advantage is that a well-constructed multiple-choice test can yield test scores at least as reliable as those produced by a constructed-response test, while also allowing for a large portion of the topics covered in a course to be assessed in a short period of time (Bacon, 2003). Well-written multiple choice items can serve to assess higher level cognitive processes, although creating such items does require more skill than writing memory-based items (Siegfried & Buckles, 2006). A lot of effort and time is required to construct good quality MCQs as compared with essay questions. This is possible if the test constructor follows rigidly the numerous guidelines for writing MCQs.

However, multiple-choice items are often criticized for focusing on what students can remember and do not assess students' abilities to apply and analyze course-related information (Walsh & Seldomridge, 2006). Another criticism is that the format of MCQs let students guess even when they have no substantive knowledge of the topic under consideration (Biggs, 1999).

However, blind guessing is quite uncommon on well-written classroom tests and in-formed guessing, which is based on a critical consideration of the question and the available options, provides a valid measure of student achievement (Downing, 2003).

To ensure the quality of a test, evidence of validity and reliability are required (Kimberlin & Winterstein, 2008). Validity is the truthfulness of a test. If a test measures what it intends to measure is called validity (Singh, 2006). Validity has its various types such as content validity, criterian-related validity, predictive validity, construct validity etc. For the present study, the researcher used content validity.

Reliability is one of the important characteristics of any test. Reliability refers to the precision or accuracy of the score (Singh, 2006). The consistency of the test is referred to as reliability. Reliability is the degree to which a measurement instrument produces consistent results when tested and re-tested as well as the degree to which a test is internally consistent (Heale & Twycross, 2015). The methods which are generally used to derive the reliability coefficient are:

(i) The same form of the test to be administered twice to the same group of individuals which is called test retest method.

(ii) Two separate but equivalent forms of the test may be administered to the same individuals which is parallel form of reliability.

(iii) The test items of a single test are subdivided into two presumably equivalent and separately scored test; the two sets of scores are correlated as though they were obtained from two equivalent forms or from two testing with the same form which is called split half reliability. (Freeman, 1962).

The researcher used test-retest method where the same form of the test is administered twice to the same group of individuals to estimate reliability coefficient.

The values of reliability coefficient should be between 0.00 and 1.00 and should not have a negative value. A test with a reliability coefficient of at least 0.70 is usually considered satisfying in terms of reliability (Fraenkel et al., 2012).

In the present study, an attempt has been made to construct and standardize a Criterion Referenced Achievement Test for Class VIII students.

Objective of the Study:

The main objective of this study is to construct and standardized Criterion Referenced Achievement Test for Class-VIII students of Elementary schools. To achieve this objective, the researcher formulated following sub objectives:

- To prepare draft of the Criterion Referenced Achievement Test for Class-VIII students of Elementary schools.
- To try out the draft of the Criterion Referenced Achievement Test for Class-VIII students of Elementary schools.
- To make item analysis of the draft of the Criterion Referenced Achievement Test for Class-VIII students of Elementary schools.
- To make selection of the final draft of the Criterion Referenced Achievement Test for Class-VIII students of Elementary schools.
- To determine validity of the Criterion Referenced Achievement Test for Class-VIII students of Elementary schools.
- To determine reliability of the Criterion Referenced Achievement Test for Class-VIII students of Elementary schools.

In the study, the researcher constructed and standardized a Criterion Referenced Achievement Test for Class-VIII students of Elementary schools using the followings steps:

i. Planning of the test.
ii. Preparation of the test.
iii. Administration of the test
iv. Item analysis
v. Standardization of the test.

The steps of construction and standardization of Criterion Referenced Achievement Test for Class-VIII students has been shown in the Fig 1. cited below:

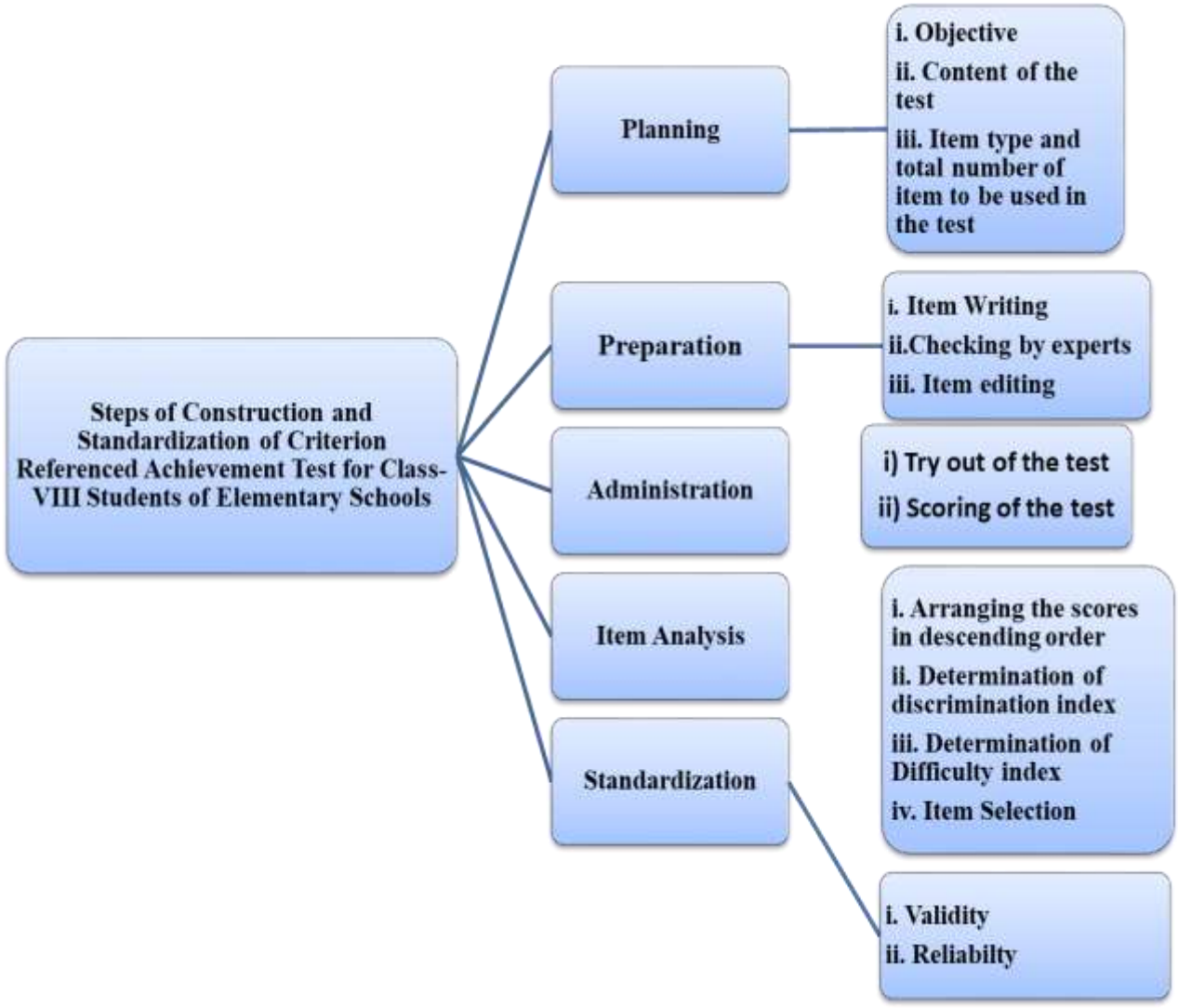

Fig I: Steps of Construction and Standardization of Criterion Referenced Achievement Test for Class-VIII Students of Elementary Schools.

Methodology

Method of the Study: Considering the nature of the objective and types of data to be collected, the researcher used descriptive survey method in the present study.

Population: The population of the study comprise of all the Class VIII students studying in Government Schools of Dibrugarh District.

Sample:
Sample for the Tryout of the Criterion Referenced Achievement Test for Class-VIII Students:

To conduct the present study, the researcher purposively selected 5 (five) Assamese Medium schools for tryout of the Criterion Referenced Achievement Test for Class-VIII Students. For try out, the test was administered to a group of comprising 200 (two hundred) students studying in Class VIII in the selected schools.

Sample for Estimating the Reliability of the Criterion Referenced Achievement Test for Class-VIII Students:

For estimating reliability, the researcher selected two Government school of Tinsukia District, using purposive sampling method. The Test-Retest reliability technique was used to estimate reliability of the test. For determining reliability, the test was administered twice to a group of comprising 110 students studying in Class VIII in the selected schools.

Construction and Standardization of the Criterion Referenced Achievement Test for Class-Viii Students:

To construct and standardize the Criterion Referenced Achievement Test for Class-VIII Students, the researcher adopted the followings steps:

i. Planning of the Test:

A Criteria Referenced Achievement Test was constructed and standardized by the researcher to measure the achievement for Class VIII students studying in Government Schools. The test was a multiple choice test based on the selected contents from all the textbooks i.e Assamese (Language I), English (Language II), Hindi (Language III), Social Science, General Science and General Mathematics of Class VIII prescribed by National Council of Educational Research and Training (NCERT).

The test items were constructed in Assamese language. A response sheet was prepared to elicit various information's of the students such as name, class, roll numbers, school name, date etc. The test contained 100 (one hundred) multiple choice items. The weightage allotted to each subject of the Criterion Referenced Achievement Test for Class-VIII Students is shown in the Table 1.

Table 1: Weightage Allotted Subject Wise to the Criterion Referenced Achievement Test for Class-VIII Students

Sl. No.	Subject	Number of Items
1	Assamese (Language I)	10
2	English (Language II)	20
3	Hindi (Language III)	10
4	Social Science	20
5	General Science	20
6	General Mathematics	20
TOTAL		**100**

The researcher selected items from all the 6 (six) subjects of Class VIII students studying in the Government schools of Assam. The weightage allotted to the domain of the Criterion Referenced Achievement Test for Class-VIII Students is shown in Table 2.

Table 2: Weightage Allotted to the Different Domains of Criterion Referenced Achievement Test for Class-VIII Students.

Sl. No	Domain	Number of Items
1	**Knowledge**	**40**
2	**Comprehension**	**35**
3	**Application**	**25**
TOTAL		**100**

The blueprint of the Criterion Referenced Achievement Test for Class-VIII Students is shown in Table 3.

Table 3: Blueprint of the Criterion Referenced Achievement Test for Class-VIII Students

Sl.No.	Subjects	Domains			Total no. of Items/Marks
		Knowledge (No. of items)	Understanding (No. of items)	Application (No. of items)	
1	English	02	14	04	20
2	Assamese	02	04	04	10
3	Hindi	05	04	01	10
4	Social Science	13	04	03	20
5	General Science	15	03	02	20
6	General Mathematics	03	06	11	20
Marks		**40**	**35**	**25**	**100**

ii. Preparation of the Test:

The researcher constructed pool of items for the present study. The researcher added more items than required to the Criterion Referenced Achievement Test draft, from which the final test would be chosen. The number of test items prepared from each lesson of all the textbooks i.e Assamese (Language I), English (Language II), Hindi (Language III), Social Science, General Science and General Mathematics of Class VIII are shown in Table 4.

Table 4: Subject Wise Lessons and Items Prepared for the Criterion Referenced Achievement Test

Subject	Lesson No.	No. of items	Total Items	Total Marks
Language I (Assamese)	1	6	40	40
	2	7		
	3	7		
	4	6		
	5	4		
	6	5		
	7	3		
	8	2		
Language II (English)	1	10	40	40
	2	10		
	3	10		
	4	10		
Language III (Hindi)	1	7	40	40
	2	7		
	3	7		
	4	7		
	5	6		
	6	6		
General Maths	1	8	40	40
	2	7		
	3	7		

After

	5	6		
	6	6		
	7	6		
General Science	1	6	40	40
	2	9		
	3	6		
	4	5		
	5	5		
	6	4		
	7	3		
	8	2		
Social Science	1	6	40	40
	2	6		
	3	6		
	9	5		
	10	5		
	11	2		
	17	2		
	18	2		
	19	2		
	20	2		
	21	1		
	22	1		

writing the items, the researcher consulted a number of times with the supervisor and subject experts to check the content coverage as well as any modification required in the test items . As per the suggestion of experts, the researcher replaced and modified some of the items to administer in the final test. The domain wise distribution of the test items in the Criterion Referenced Achievement Test for Class-VIII Students is shown in Table 5.

Table 5: Domain Wise Distribution of the Test Items in the Criterion Referenced Achievement Test for Class-VIII Students

Sl.No.	Subject	Knowledge	Comprehension	Application	Total
1	Assamese (Language I)	1,2,3,4,5,7,9,16,17,19,20,22,23,24,29,30,31,32,33,35,36,38	8,11,12,13,14,15,18,21, 25,27,28, 34,37	6,10,26,39,40	40
2	English (Language II)	2,3,4,6,11,12,13,15,17,19,20,22,28,29,30,33,34,36,39	1,5,7,8,9,10,16,18,21,24,25,26,31,32,35,37,38	14,23,27,40	40
3	Hindi (Language III)	1,2,3,4,5,9,12,14,15,16,18,22,24, 25,26,28,29,30,31,32,34,37,38,39	6,7,8,10,11,13,17,19,20 ,21,23,27,33,36	35,40	40
4	Social Science	4,5,6,7,8,9,10,11,12,13,16,18,19, 21,22,25,26,27, 29,30,31,35,37,39,40	1,2,3,14,15,17,24, 28,32,34	20,23,33,36,38	40
5	General Science	1,4,5,6,7,8,9,13,15,16,17,19,20,21,26,28,30,31,32,33, 37,39	2,3,10,11,12,14,18,22,23,25,27,29,35,36,38	24,34,40	40
6	General Mathematics	1,9,19, 24,27,34,39	3,4,5,7,16,17, 21,22,25, 32,36,40	2,6,8,10,11,12,13,14,15,18,20,23,26,28,29,31,31,33, 35,37,38	40

iii. Administration of the Test:

For administration the Criterion Referenced Achievement Test and try out of the draft of the test, purposive sampling technique was used. For try out, the test was administered to a group of comprising 200 (two hundred) students studying in Class VIII in the selected schools.

(a) Instruction: The researcher prepared necessary instructions and attached to the front page of the draft test. Oral instructions were also given.

(b) Scoring Procedure: The process of scoring in the Criterion Referenced Achievement Test for Class-VIII Students is shown in Table 7.

Table 7: Scoring of the Criterion Referenced Achievement Test for Class-VIII Students

Answer	Marks
Correct	1
Incorrect	0

iv. Item Analysis:

After scoring of the test, the researcher has arranged the scores of all the sampled 200 students in descending order, i.e., from the highest to the lowest score. The top 27% and the bottom 27% of the students were selected based on the total score obtained in the test to form two separate groups namely, high group and low group.

Then, the discrimination index (V) and difficulty index (P) value for each test item were calculated to find out whether the discrimination index and difficulty index of a particular test was significant or not. The discrimination index (V) of each item was calculated using the following formula:-

$$V = \frac{R_U}{N_U} - \frac{R_L}{N_L}$$

In the formula, V represents the net discrimination index of each item, R_U represents the number of students giving responses correctly in the upper group, R_L represents the number of students giving responses correctly in the lower group, N_U represents the number of students in the upper group, N_L represents the number of students in the lower group. Items with a discrimination index of more than 0.36 (Ebel & Frisbie,1991) were kept for calculation of reliability.

After calculating discrimination index, the researcher calculated difficulty index (P) of each item by using the following formula:-

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

Where, P represent the net difficulty index of each item, R_U represents the number of students giving responses correctly in the upper group, R_L represents the number of students giving responses correctly in the lower group, N_U represents the number of students in the upper group, N_L represents the number of students in the lower group. Item with a difficulty index between 0.36 to 0.70 (Ebel & Frisbie, 1991) were kept in mind for selection of the items.

The distribution of the test items having satisfactory discrimination index and difficulty index are shown in Table 8.

Table 8: Distribution of the test items having satisfactory discrimination index and difficulty index.

Sl. No.	Subject	Knowledge (Sl. No. of test items)	Comprehension (Sl. No. of test items)	Application (Sl. No. of test items)	Total No. of Test Items
1	English	12,34	1,7,8,9,10,16,18,21,26,31,32,35,37,38	14,23,27,40	20
2	Assamese	4,24	12,13,14,15	10,26,39,40	10
3	Hindi	2,12,13,22,26	6, 7,20,35	40	10
4	Social Science	6,7,9,10,13,21,22,25,29,31,35,37,40	2,3,14,15	20,33,38	20
5	General Science	1,5,7,8,9,13,15,20,26,28,30,31,33,37,39	3,10,11	24,34	20
6	General Mathematics	9,24,34	3, 7,22,25,32,36	2,8,10,11,12,13,14,20, 23,26,31	20
Total		**40**	**35**	**25**	**100**

On the basis of satisfactory discrimination index and difficulty index, the researcher decided to include these 100 items for the final test. Subject wise selected items for final test are shown in Table 9.

Table 9: Subject Wise Selected Items for the Final Test

Sl.No	Subjects	DOMAIN			No. of items selected
		Knowledge	Comprehension	Application	
1	English	02	14	04	20
2	Assamese	02	04	04	10
3	Hindi	05	04	01	10
4	Social Science	13	04	03	20
5	General Science	15	03	02	20
6	General Mathematics	03	06	11	20
TOTAL		**40**	**35**	**25**	**100**

v. Standardization of the Test:

Validity of the Test:

To determine the content validity, the draft Criterion Referenced Achievement Test for Class VIII Students was sent to a number of experts to check whether the blueprint covers all the content of the test. Based on their suggestions, necessary modifications were done in certain test items.

Reliability of the Test:

Test-Retest technique was used to estimate the reliability of the Criterion Referenced Achievement Test for Class VIII Students. For determining reliability, the test was administered twice to a group of comprising 110 students studying in Class VIII in the selected schools within a gap of 10 days. Two sets of score were obtained from these tests. Then the researcher used Pearson Product-Moment Correlation method to calculate the reliability coefficient. The reliability coefficient of the test was found as 0.77.

Conclusion

The study was carried out to construct and standardize the Criterion Referenced Achievement Test for Class VIII students studying in Government Schools.

The test items were constructed in Assamese language. The test contained 240 (two hundred and forty) multiple choice items. The researcher selected 100 test items from 240 test items for the final test. The test was constructed and standardized on the sample of 310 Class VIII students studying in Government Schools. Content validity was estimated and Test-Retest technique was used to determine reliability of the test. The reliability coefficient of the test was found as 0.77 which is quite satisfactory.

References:

Anastasi, A., & Urbina, S. (2019). *Psychological testing* (7th ed.). Pearson.

Apino, E., & Retnawati, H. (2017). Developing instructional design to improve mathematical higher order thinking skills of students. *Journal of Physics: Conference Series*, 812(1). https://doi.org/10.1088/1742-6596/812/1/012100

Bacon, D.R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education*. 25(1), 31-36. https://doi.org/10.1177/0273475302250570

Baruah, P., Gogoi, M. (2022). A construction and standardization of a concept Achievement Test in mathematics for class-Viii students. *Child Studies in Asia-Pacific Context (CSAC)*. 12(1), 184-191 https://www.researchgate.net/publication/364735881

Biggs, J. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research and Development*, *31*(1), 39–55. https://doi.org/10.1080/07294360.2012.642839

Bolt, S. (2011). Making consistent judgments: Assessing student attainment of systemic achievement targets. *The Educational Forum*. 75(2), 157–172. https://doi.org/10.1080/00131725.2011.552694

Bond, L. A. (1996). Norm and criterion-referenced testing. *Practical Assessment, Research, and Evaluation*. 5(1), 1-3, https://doi.org/10.7275/dy7r-2x18

Bracey, G. (2000*). Thinking about tests and testing: A short primer in assessment literacy.* American Youth Policy Forum.

Brown, S. (1998). Criterion-referenced assessment: What role for research. In H. Black & W. Dockerell (Ed.), *New developments in educational assessment* (3rd ed., pp. 01-14). Monograph Series

Chaudhary, S., & Tyagi, P. S. K. (2017). Construction and standardization of Achievement Test in educational psychology. *Educational Quest: An Int. J. of Education and Applied Social Science*, *8*(3), 817–823. https://doi.org/10.5958/2230-7311.2017.00141.6

Cullari, S. (1998). *Foundations of clinical psychology*. Allyn & Bacon

Dickson, A., Jephtar, A.M., & Dennis, A.D. (2020). Test, measurement and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education*, 9(1), 109-119. https:dx.doi.org/10.11591/ijere.v9i1.20457

Downing, S. M. (2003). Validity: On meaningful interpretation of assessment data. *Med Educ*. 37(9), 830-837, 10.1046/j.1365-2923.2003.01594.x. PMID: 14506816

Ebel, R.L. & Frisbie, D.A. (1986).*Essentials of educational measurement*. Englewood Cliffs. Prentice Hall.

Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement*. Prentice Hall of India Private Limited.

Facione, P. A.,Facione, N.C., & Giancarlo, C.A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, *20*(1). https://doi.org/10.22329/il.v20i1.2254

Fraenkel, J. R., Wallen, N. E. & Hyun, H. H. (2012). *How to design and evaluate research in education*. The McGrow-Hill Companies. https://saochhengpheng.files.wordpress.com

Freeman, F. S. (1962). *Theory and practice of psychological testing*. Oxford and IBH Publishing Co. Pvt. Ltd.

Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple choice questions (MCQS) from an assessment of medical students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine*, *39*(1), 17–20. https://doi.org/10.4103/0970-0218.126347

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*(8), 519–521. https://doi.org/10.1037/h0049294

Gogoi, B., & Bhuyan, S. (2023). Construction and standardization of an Achievement Test in mathematics and english grammar for class ix students. *International Journal of Creative Research Thoughts* .11(7), 98-115. https://ijcrt.org/papers/IJCRT2307131.pdf

Heale, R., & Twycross, A . (2015). Validity and reliability in quantitative studies. *Evid Based Nurs*, 18(3), 66-67. https://ebn.bmj.com/content/ebnurs/18/3/66.full.pdf

Kimberlin, C.L., & Winterstein, A.G. (2008) Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(23), 2276–2284. https://doi.org/10.2146/ajhp070364

Mamolo, L. A. (2021). Development of an Achievement Test to measure students competency in general mathematics. *Anatolian Journal of Education*. *6*(1), 79-90. https://doi.org/10.29333/aje.2021.616a

Odinaka, P., Victor, A. V., Ossai, O. V., Ugwoezuonu, A. U., Egenti, N. T., & Eseadi, C. (2017). Procedures for the construction and development of psychological test. *European Journal of Scientific Research*. 145(2). http://www.europeanjournalofscientificresearch.com

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1). https://doi.org/10.1080/2331186X.2017.1301013

Siegfried, J & Buckles, S. (2006). Using multiple-choice questions to evaluate in-depth learning of economics. *Journal of Economic Education*. 37(1). 48-5. https://www.jstor.org/stable/30042686

Singh, T., Gupta, P.,& Singh, D. (2009) . *Test and item analysis*. In: Principles of Medical Education. Jaypee Brothers Medical Publishers (P) Ltd.

Singh, A. K. (2006). *Tests, measurements and research methods in behavioural sciences*. Bharati Bhawan Publishers & Distributors.

Walsh, C. M., & Seldomridge, L. A. (2006). Critical thinking: Back to square two. *Nursing Education*, 45(6), 212-219. https://doi.org/10.3928/01484834-20060601-05