

CONSTRUCTION AND STANDARDIZATION OF ACHIEVEMENT TEST IN ENGLISH

Dr. Ram Mehar*, Dr. Sonia Rani**

*Associate Professor, Department of Education, USOL, Panjab University, Chandigarh

**TGT (English) Kendriya Vidyalaya No.4, Ambala Cantt.

ABSTRACT

This paper highlights the process of development and standardization of achievement test in English for Class X, Central Board of Secondary Education, New Delhi. Initially a preliminary draft of criterion referenced test consisting of 120 items was prepared. Eight subject experts were personally requested to reflect their opinion upon every statement. After getting feedback from themselves; some items were modified and 35 items were identified as dead distracters. A total of 85 test items were preserved for the final draft of Criterion Referenced Test in English. The tryout of criterion referenced test was taken to a sample of 50 students of class 10th studying in Kendriya Vidyalaya, Central Reserve Police Force, Pinjore. A criteria of Kelley (1939) method was opted for item analysis i.e. to find out Difficulty Value (DV) and Discrimination Power (DP) of the criterion referenced test. After try out and item analysis of criterion reference test, 70 items were retained for the first draft of achievement test in English. Traditional way of investigating reliability and validity has been criticized as inappropriate for criterion test (Popham & Husek, 1969). Therefore, alternative analysis for investigating the adequacy of criterion test has been developed (Berk, 1980). In contrast to the correlation statistics, these analyses rely minimally on the notion that inter-individual variability is necessary (Carver, 1974; Hambleton & Novick, 1973; Huynh, 1976). Popham (1975) suggested less sophisticated but more meaningful reliability measures to be employed when marked range restrictions are present. For reliability measures he had suggested few techniques. The investigator selected the one in which percentage of scores of students were to be calculated on two different occasions. The development of achievement test passed through three stages: (i) first draft of achievement test (ii) second draft of achievement test (iii) final draft of achievement test. So, the first draft of achievement test consisted of 70 items. After try out of first draft of achievement test to 50 students of class 10th studying in Kendriya Vidyalaya, Central Reserve Police Force, Pinjore; the same criteria was adopted for item analysis i.e. to find out Difficulty Value (DV) and Discrimination Power (DP) for the first draft of Achievement Test. A total of 60 items were

remained for the second draft of achievement test in English. After try out upon 75 students of class 10th studying in Kendriya Vidyalaya, No 1,2 and 3 Ambala Cantt., the above same criteria was adopted for item analysis for the second draft of achievement tes in English. After item analysis of second draft of achievement test in English, a total of 50 items were remained left for the final draft of achievement test in English. The reliability of final draft of achievement test was conducted on sample of 50 students of class 10th studying in Kendriya Vidyalaya, No 1, 2 and 3 Ambala Cantt. The test-retest reliability of the measure of achievement test was found to be 0.80. The Validity was established by content validity method.

Keywords: Construction, Standardization, Criterion Referenced Test, Achievement test in English

Introduction

An achievement test is a controlling process for teaching learning activity. *Achievement tests* help in evaluating the effectiveness of teaching instructions. Achievement tests are past and present oriented which register the degree of learning or achievement after instructions. According to Dowine and Heath (1974) any test that measures the attainments or accomplishments of an individual after a period of learning is called an achievement test. It can also be thought of as a sample of indicator of a student's knowledge taken at a particular time. Therefore, achievement tests propose to measure what and how much pupils have learnt as a result of formal and informal instructions. The achievement tests differ from intelligence or aptitude tests in that the former measures the quantity or quality of learning attained in a subject of study or group of subjects after a certain period of instruction and the later measures pupil's innate capacity for attainment or accomplishment independent of learning. Lindeman (1967) classified achievement tests in following categories such as (i) Teacher Made Test (ii) Standardized Test.

Teacher made tests are frequently the basis of evaluating a student's progress in school. These are more specifically focused and they usually reflect the content of a particular unit or course. On the other hand, standardized tests are particular kinds of tests, different from the final examination: a high school teacher might design for his course. When talking about tests, "standardized" simply means that everyone who takes the test is given the same amount of time and sees the same or very similar test questions. These tests are widely used because, by and large, they have shown to be an efficient way to collect information about what people know and can do. For the present study, the investigator could not lay hands on appropriate standardized achievement test in English, on the selected topics. So a need was felt to develop one, so as to evaluate the outcome of instructional strategies, an achievement test was developed from

criterion referenced tests, that exhibited significant response variant to measure the performance of students.

Development and Standardization of Criterion Referenced Test in English

A criterion-referenced test is a test that provides a basis for determining a candidate's level of knowledge and skills in relation to a well-defined domain of content. Criterion-referenced tests, a type of test introduced by Popham and Husek (1969), is also known as domain-referenced tests. Glaser (1963) proposed that a criterion-referenced measure was related to a student's acquisition of knowledge along a continuum ranging from no proficiency to perfect performance. Kriewall (1972) indicated that a criterion referenced test is one in which items are homogeneous in difficulty for each examinee. A criterion referenced test is not only having items of equivalent difficulty but equality in form and content as well by Emrick (1971). According to Ivens (1970) a criterion referenced test is a one composed of items keyed to a set of behavioural objectives. Livingston (1972) defined a criterion referenced test as one for which a criterion score is specified without reference to the distribution of the scores in the group. Pophem (1975) suggested the following steps for the construction of criterion referenced test such as: (i) Domain definitions (ii) Generating items (iii) Improving items (iv) Reliability and Validity.

- (i) *Domain Definitions*- In this phase the limits of behavior that the test items would measure and to which all individual performance is referenced is determined. Next aspect regarding domain definition is selection among competing domain alternatives, which is affected by transferability within domain-alternatives and transferability outside the domain. After identification and definitions of learning outcomes, units are selected from the prescribed work book.
- (ii) *Generating Items*- A number of items may be constructed for any given objective. In terms of feasibility, a survey of current measures revealed that the usual practice is to use about 3 to 5 items per objective. As opined by Gronlund (1977) each item should be started as multiple choice items and construct to other types of items when the learning outcomes of the subject matter made is desirable to do so. In writing the items proper care should be taken regarding the coverage of the content structure of item type and language. After generating items, each type of items should be grouped together and appropriate directions for the students to attempt the items should be written. The scoring key should also be developed. Then the preliminary draft of criterion referenced test will be developed.
- (iii) *Improving Items*- A try out should be conducted to ensure that the entire class of learner's behavior has been circumscribed. The preliminary draft should be administered to a sample of students of prescribed class who has already studied the content. No time limit

should be imposed. The answer-sheet should be scored with the help of scoring key prepared for the test.

- (iv) *Reliability and Validity*- The idea of reliability is crucial for criterion referenced test as for all other tests. Before administering the final test reliability should be calculated by using any method i.e. test-retest, split-half etc. likewise validity should also be calculated.

Content Specification

Pertaining to the content and objectives selected for the study the investigator felt a need to develop a criterion referenced test to evaluate the student's achievement on the content selected for the study. The details of the selected content have been given in table-1.

Table 1: Details of the selected content

No.	Units
1	<i>Determiners</i> : article, demonstrative , possessive, interrogative
2	<i>Tenses</i> : The Present Tense, The Past Tense, The Future Tense
3	<i>Subject- Verb Agreement</i>
4	<i>Active-Passive Voice</i> :Use in Present Tense, Use in Past Tense, Use in Future Tense
5	<i>Reported Speech</i> : Direct Speech , In-direct speech

Table-1 shows that five units of English grammar of class X were selected from the prescribed syllabus of Central Board of Secondary Education, New Delhi and a few questions were prepared on each topic.

Specification of Instructional Objectives

Instructional objectives state the expected performance level of the students in specific, observable behavioral terms under defined conditions at designated time. In any instructional system design preparation of the statement of instructional objectives is the most critical and important step as it helps the designer to design the instructional programme. After analysis of the content, the 60 Instructional objectives were specified. The instructional material under knowledge, comprehension and application category of objectives was identified.

Development and Standardization of Criterion Referenced Test in English

The development of a good criterion referenced test in English involves (i) Planning (ii) Writing and editing of test items (iii) Try-out and item analysis (iv) Reliability and Validity.

- (i) *Planning of the Test*: Mehar (1997) in his thesis reported Stanley (1964) indicating that planning the test consists of the advance decisions about the unit of content, instructional

objectives, the type of test items to be used in the test and weightage to be given to various aspects. After identification and definition of instructional objectives, test items to be used in the test corresponding to the objectives were constructed. For testing 60 objectives a total of 120 test items were planned. After studying thoroughly the syllabus prescribed for X English students from the prescribed syllabus of Central Board of Secondary Education, New Delhi. It was also decided that test was meant for students of age 13 to 18 years. The main content area was identified and major course instructional behavioral objectives were specified. Twenty lessons based on interactive whiteboard technology from English grammar from class X were selected for present study. The investigator developed multiple choice questions; fill in the blank and true/false type questions keeping in mind the content and objectives of instructional material. For the present test, a blue print with three dimensions viz. knowledge of content, understanding and application of content was prepared. The specification of criterion test has been given in table-2:

Table 2: Specification of the criterion referenced test in English

No.	Units	Objectives	Type of Test Items						Total
			Knowledge		Understanding		Application		
			F	MC	F	MC	F	MC	
1	Determiners	08	07	----	----	09	-----	10	26
2	Tenses	27	----	10	----	06	----	08	24
3	Subject-Verb Agreement	02	----	05	----	07	----	03	15
4	Active-Passive Voice	12	----	12	----	11	----	12	35
5	Reported Speech	11	----	07	----	06	----	07	20
	Total	60	07	34	----	39	----	40	120

Here: F signifies Fill in the Blank and MC signifies Multiple Choice item

Table-2 shows that 8 objectives were prepared for the first unit, 27 objectives were formed for second unit, 2 objectives were prepared for third unit, 12 objectives were formed for the fourth unit and 11 were prepared for fifth unit. The items were prepared to be of fill in the blanks type, items were made in the form of true false type and items were planned of multiple choice questions. Hence, the total 120 items and 60 objectives were framed for the preliminary draft of criterion referenced test

(ii) *Writing and Editing of Test Items-* Keeping in view the quality of test items, the investigator constructed items for the criterion referenced test. The test items were written in a simple and clear language. Non-functional words were not included in the items as they tend to lower the validity of the item. The exceptionally long term and double barreled items were avoided. After generating items, each type of items were grouped together and appropriate directions for the students to attempt these items should be clearly written in the

test. As opined by Gronlund (1977) each item should be started as multiple choice items and construct other types of items when the learning outcomes of the subject matter made it desirable to do so. In writing the items, proper care should be taken regarding the coverage of the content structure of the items type and language. The item of each unit should be exhaustively written so as to test every aspect thoroughly. This test was given to subject experts/ teachers from school, college and university to check the items for content and language, to check relationship between objectives of content and test items. The investigator with her supervisor devoted several sittings to consider the judgement of the experts on the statement. Discussion with subject experts was held individually. After obtaining feedback from experts, some test items were modified and 35 test items i.e. 1, 2, 3, 7, 12, 13, 19, 25, 27, 28, 41, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 101, 102, 103, 104, 105, 107, 110, 111, 112, 113, 114 and 115 were identified as dead distractors. The 85 test items were retained for the final draft of the criterion referenced test. The items were rearranged in logical order.

(iii) Try-Out and Item Analysis: The final draft of the criterion referenced test was cyclostyled and required numbers of copies were prepared. The final draft criterion referenced test in English was administered to a sample of 50 students drawn from 10th class of Kendriya Vidyalaya, Central Reserve Police Force, Pinjore, who had already studied the selected content. No time limit was imposed but it was found that the majority of the students completed the test in about one hour. The observations made by the students were noted down and considered for revising the draft. It may be mentioned here that unlike the other tests, the criterion test is to describe the entry level of the students and the extent to which the learners have attained the pre-determined set of terminal behavior. Shaycoft (1979) also stressed that the item analysis is not nearly essential for criterion test. But even for the criterion tests that have to be constructed in such a way that items cannot be thrown out after try-out, item analysis may nevertheless prove useful to identify items or options that seem not be working well (Meher 1997, p.56). The responses of 85 multiple choice items were analyzed and distractors were revised in such a way that they fell in close proximity of the correct response region and no distracter was made so attractive as to compete with the correct response. Only 15 items i.e. 23, 34, 35, 38, 39, 57, 59, 60, 61, 65, 66, 69, 71, 76, 77 were identified as dead distractors and they were removed from final draft of the criterion referenced test. Thus, 70 items were retained for the first draft of the achievement test.

(iv) Reliability and Validity: Traditional way of investigating reliability and validity has been criticized as inappropriate for criterion test (Popham & Husek, 1969). Therefore, alternative analysis for investigating the adequacy of criterion test has been developed (Berk, 1980). In contrast to the correlation statistics, these analyses rely minimally on the notion that

inter--individual variability is necessary (Carver, 1970; Hambleton & Novick, 1973; Huynh, 1976). Popham (1975) suggested less sophisticated but more meaningful reliability measures to be employed when marked range restrictions are present. For reliability measures he had suggested few techniques. The investigator selected the one in which percentage of scores of students were to be calculated on two different occasions. The percentage of students corresponding to two different scores classes has been presented in table-3:

Table 3: Percentage of students corresponding to different classes of scores

Difference of Scores	Percentage of Students
0 -5	33
6-10	32
11-15	18
16-20	7
21-25	5
26-30	2
31-35	1
36 and above	2
Total	100

Table-3 shows that the majority of the proportion of differences between two student's scores falls in first two categories. This percentage decreases rapidly with increasing score classes. It depicts that the students show a marked level of consistency across the scores level. So, this test may be considered reliable for measuring of performance of students.

The test was validated against the criterion of content validity. It was the only kind of validation which was usually feasible under the circumstances. Gronlund (1977) also opined that the, type of validity of greatest importance for criterion referenced test is content validity. The content validity of the test was established by relating the task to the instructional objectives. The correspondence between the two was determined. Thus, 70 items were retained for the first draft of the achievement test in English grammar.

Achievement Test in English Grammar

It has been recognized that achievement test is an integral and important part of educational process. It measures the extent to which a person has achieved acquired certain information of mastered certain skills, usually as a result of specific instruction (Stanley & Hopkins, 1972). Ebel (1966) was of the view that achievement test is a sample of indicator of a student's knowledge taken at a particular point of time. An achievement test can be designed for two purposes. First, performance can be measured to provide information about the

characteristics of students' present behavior. Second, achievement can be measured to provide information about the instructional treatment which produces that behaviour. So from the above point of view the development of an achievement test was a crying need. Achievement test was developed from those criterion test items which showed significant response variance to measure the performance of students. It can be differentiated from the criterion test on the basis of purpose, difficulty range, normality of conditions etc. Development of an achievement test passed through three stages such as (i) first draft of achievement test (ii) second draft of achievement test (iii) final draft of achievement test.

First Draft of Achievement Test in English Grammar

The first draft of the achievement test was prepared on basis of the 70 items selected after the validation of criterion referenced test. The first draft consisted of these items which were accepted as such and those which were modified taking into consideration the discriminating power and difficulty value. Development of achievement test involves following steps (i) Planning (ii) Preparation (iii) Try out and Evaluation. The planning and preparation of the test items have already been discussed in the previous section. Now, the investigator will discuss tryout and evaluation.

- *First Tryout and Evaluation:* The items of the criterion test were used as items of first draft of achievement test and were administered to a sample of 50 students drawn from the 10th class of Kendriya Vidyalaya, Central Reserve Police Force, Pinjore who had already covered the content of the test. Normal testing conditions were ensured to the students. After the test was completed by all the students, the papers were collected and scored with help of scoring key developed by the investigator. Each item response marked correctly by the students was given one mark. No correction for guessing was applied as very negligible percentage of students made omissions. Stanley and Hopkins (1972) along with Ebel (1966) have suggested the application of correction for guessing when some students have omitted a fairly large number of items. After the scoring, the item-analysis of the test was carried out. The items were analyzed qualitatively, in terms of their content, form and quantitatively in term of their statistical properties. For this all the scored answer sheets were selected with highest scores and the same number with lowest scores were selected to form upper and lower groups (Kelly, 1939). For the calculation of discriminatory power and difficulty value the following procedure was followed:

- (i) The answer sheets of all the students were arranged in descending order.
- (ii) The top 27% forms the upper group and bottom 27% formed lower group.
- (iii) After that the correct responses for each item in both the groups were calculated.

Each group consisted of 13 students. As such the difficulty value and discriminating power were calculated from those sub groups making a total of 26 students. For calculating difficulty value and discrimination power, the following formulas were used:

$$\text{Difficulty Value} = \frac{R_U + R_L}{N} \qquad \text{Discriminating Power} = \frac{R_U - R_L}{N/2}$$

Where:

R_U = Number of right responses in the upper group

R_L = Number of right responses in the lower group

N = Total number of students in both the groups

In this way difficulty value (DV) and discriminating power (DP) for each item of first draft of achievement test in English were computed. Each group consisted of 13 students. The difficulty value and discriminating power were calculated from those sub groups making a total of 26 students. The index of difficulty value (DV) and discriminating power (DP) for each item were computed. The distribution of discriminating power of items of the first draft of achievement test in English has been given in table-4

Table 4: Distribution of discriminating power of items of the first draft of achievement test

S N	DP	f	Item Number	Remarks
1	0.40 and above	35	3,6,10,11,12,14,16,17,18,19,22,23,24,25,26,28,29,30,32,33,35,36,40,41,43,46,54,55,58,61,62, 64, 65, 66, 69.	Very good items
2	0.30 to 0.39	16	1,5,8,9,21,27, 31, 34, 39, 44, 48,49,57,60,63,68.	Reasonably good items
3	0.20 to 0.29	9	2,13,15,47,53,56,59,67,70.	Marginal items
4	Below 0.19	10	4,7,20,37,38,42,45,50,51,52	Poor items

Table-4 shows that the discrimination power of the first draft of achievement test was analysed that out of 70 items, the 35 items were considered very good and needed no revision when the discrimination power is 0.40 and above. Accept, scrutinize for improvement of item when 16 items with discrimination power is between 0.30 and 0.39 were considered reasonably good items. 9 items with discrimination power between 0.20 and 0.29 were regarded as marginal items while 10 items having discrimination power 0.19 and below it deserved to be eliminated. The discriminating power of the test items of final draft of criterion referenced test was ranged from 0.07 to 0.69. The distribution of difficulty value of items of the first draft of achievement test in English has been given in table-5.

Table 5: Distribution of difficulty value of items of the first draft of achievement test

S N	DV	f	Item Number	Remarks
1.	Above 0.75	10	4,7,20,37,38,42,45,50,51,52.	Easy items
2.	0.50 to 0.75	60	1,2,3,5,6,8,9,10,11,12,13,14,15,16,17,18,19,21, 22,23,24,25,26,27,28, 29,30,31,32,33,34,35,36,39, 40,41,43,44,46,47,48,49,53,54,55,56,57,58,59,60, 61,62,63,64,65,66,67,68,69,70,	Marginal items
3.	0.25 to 0.49	0	----	Reasonably good items
4.	Below 0.25	0	----	Difficult items

The table-5 shows that the difficulty value of the first draft of achievement test was analysed that out of 70 items, there were 10 items having difficulty value above 0.75 indicating that these were very easy items, 60 items having difficulty value between 0.50 to 0.75 were regarded as marginal items, zero items were regarded as reasonably good items while zero item having difficulty value below 0.25 were considered to be difficult items. The total 10 easy and difficult items were eliminated from the final draft of criterion referenced test in English. The major reason for measuring item difficulty was to choose items of suitable difficulty level. The criteria given by Ebel (1966) say that the items having difficulty level above 0.75 and below 0.25 were rejected as they were very easy and very difficult items respectively. The items having difficulty value ranged from 0.25 to 0.75 were accepted. However, the difficult level of items for the first draft ranged from 0.54 to 0.96. The second draft of achievement test in English comprised of 60 items. These items were analyzed and evaluated to enhance the quality of items. The distracters which contend with the right answer can be made less alluring to maintain a strategic distance from a more noteworthy interest to the students when contrasted with that of the right answer. The distribution of distractors competing with correct answer and weak distractors in the first draft of the achievement test has been given in table-6.

Table 6: The distribution of distractors competing with correct response and weak distractors of the first draft of achievement test in English

S.No.	Form of Response	Item No	f
1	Distractors competing with correct response	28, 29, 30, 31	4
2	Weak Distractor	4, 7, 20, 37, 38, 42, 45, 50, 51, 52	10

Table-6 shows that out of the total 70 multiple choice items only in case of 4 items distractors competing with the correct answer were modified or revised taking into consideration. The 10 items with poor distractors were removed from the first draft. So, finally the remaining 60 items

were retained. The selected and rejected items of the test after calculation of discrimination value and discriminating power have been given in table-7.

Table 7: Selected and rejected items for the second draft of achievement test in English

S. No.	f	Item no.	Remarks
1	60	1,2,3,5,6,8,9,10,11,12,13,14,15,16,17,18,19,21,22,23, 24, 25, 26,27, 28,29,30,31,32,33,34,35,36,39,40,41,43,44,46,47,48,49,53,54,55,56, 57,58,59,60,61,62,63,64, 65,66,67,68,69,70.	Selected items
2	10	4,7,20,37,38,42,45,50,51,52.	Rejected items

Table-7 shows that 60 items out of the total 70 items were retained while a total of 10 were rejected. So, finally the remaining 60 items were retained for the second draft of achievement test.

Second Draft of Achievement Test in English Grammar

The second draft of achievement test in English was prepared on the basis of the item analysis of multiple choice items. The 60 items of second draft of achievement test in English consisted of those items which were accepted as such and which were modified or revised taking into consideration the discriminating power and the difficulty value.

- *Second Try-out and Evaluation-* For the tryout of the second draft of the achievement test, a sample of 75 students was selected randomly from the 10th class students of three schools such as: Kendriya Vidyalaya No. 1,2 &3 Ambala Cantt., who have already studied the content for the second tryout. The required numbers of copies of this draft were obtained. 25 students were selected from each of the three schools for second tryout of the achievement test. An attempt was made to provide uniformity in administration and excellent testing conditions. All answer sheets returned, were scored with the help of the scoring key. After scoring the answer-sheets, the item-analysis of second draft of achievement test was carried out. The discriminating power, difficulty value and responses to distracters were calculated. The index of discriminating power and difficulty value along with the numbers belonging to each category has been recorded in table-8.

Table 8: Distribution of discriminating power of items for the second draft

No.	D.P.	f	Item No.	Remarks
1	0.40 and above	33	2, 3, 8, 10, 12, 15, 17, 19, 21, 22, 24, 26, 27, 34, 36, 37, 39, 40, 43, 45, 47, 48, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	Very good items
2	0.30-0.39	10	1, 4, 5, 9, 13, 14, 25, 28, 44, 46	Reasonably good items
3	0.20-0.29	07	11, 16, 20, 23, 32, 33, 49	Marginal items
4	0.19 and below	10	6, 7, 18, 29, 30, 31, 35, 38, 41, 42	Poor items

Table-8 shows that 33 items having discriminating power more than 0.40 were considered as very good items, 10 items with D.P between 0.30 to 0.39 were considered considerably good items, 07 items with value between 0.20 to 0.29 were regarded as marginal items while 10 items having discriminating power between 0.19 and below deserving to be eliminated. The discriminating power of the test items of the second draft ranged from 0.07 to 0.62. The index of reasonably difficulty value along with item number of every category has been given in table-9.

Table 9: Distribution of difficulty value of items of the second draft of achievement test

No	D.V.	f	Item No.	Remarks
1	Above 0.75	4	6, 7, 18, 29	Easy items
2	0.50—0.75	35	1, 2, 3, 4, 5, 8, 11, 12, 13, 14, 17, 19, 22, 24, 25, 26, 27, 36, 37, 39, 40, 43, 46, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	Good items
3	0.25—0.49	15	9, 10, 15, 16, 20, 21, 23, 28, 32, 33, 34, 44, 45, 47, 48	Marginal items
4	0.19and below	6	30, 31, 35, 38, 41, 42	Difficult items

The table-9 shows that there are 4 items having discriminating value above 0.75 indicating that these are very easy items and about 6 items having discriminating value below 0.19 are designated as very difficult ones. These 10 items were eliminated from the second draft. The difficult level of the second draft ranged from 0.54 to 0.96. There were 50 items with value ranging from 0.25 to 0.75 which were retained for the final draft of the achievement test.

Final Draft of Achievement Test in English Grammar

For the selection of items be included in the final draft of the achievement test in English grammar. The criteria given by Dececco and Crawford (1974) was given due consideration which has been described earlier. After calculation of difficulty values and discriminatory power

of all the items, 50 items were retained for the final draft of the achievement test. The most easy and most difficult items were rejected. These also included items with poor discriminatory power. The selected and rejected items for the final draft of the achievement test have been given in table-10.

Table 10: The Selected and rejected Items for the final draft

No.	f	Item No	Remarks
1	50	1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 32, 33, 34, 36, 37, 39, 40, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60.	Selected items
2	10	6, 7, 18, 29, 30, 31, 35, 38, 41, 42.	Rejected items

Table-10 indicates that 50 items out of the total 60 multiple choices items were selected to be included in the final draft of achievement test and 10 items were rejected from the second draft of achievement test in English. So, the final draft of the achievement test consisted of 50 items which included 45 fill in the blanks and 5 tick the right option type items, which all are in the form of multiple choice type items.

Administration

The final draft of achievement test in English grammar was designed to administer to the individual performance of the students in English grammar. After the distribution of the test, the subjects are required to fill in the identifying information on the cover page. They are told to read the instruction. Supervision is needed and additional explanation may be given to ensure that subjects have understood the test directions before proceeding. The test has the time limit of 45 minutes. They were told not to turn the page until they are told to do so. Writing on test paper is not allowed.

Scoring

Scoring key was made and got examined. The appropriate response sheets were assessed with the assistance of the scoring key. For each right reaction, one mark was appointed and there was no negative stamping for any wrong reaction made by the students. Total marks of the test were 50 marks.

Reliability

Reliability concerns the extent to which measurement is repeatable, i.e. when different people make the measurement on different occasions, with supposedly alternative instruments for

measuring the same thing (Nunnally, 1982). Reliability is essential to the effectiveness of any data gathering procedure (Best & Kahn, 1989). Guilford (1967) defines reliability as the proportion of true variance in the observed test scores. As the achievement test in English being heterogeneous and test items having being arranged logically, the two halves could not have been identical. Therefore, test- retest method of reliability was found to be most suitable for the test. Mouly (1970) remarked, "The test-retest method is the only feasible approach to the establishment of the reliability of the test. In other words, measurement is intended to be stable over a variety of conditions in which essentially the same results should be obtained. For determining the reliability of the achievement test, it was administered to 50 students of Kendriya Vidyalaya, Ambala Cantt at two different occasions.. The second administration of the test was given after two weeks. The test-retest product moment coefficient of correlation for the two scores was computed. The coefficient of correlation between two test scores was found to be 0.80. This coefficient of correlation is fairly high, which testifies of the soundness of the test. Therefore, the achievement test may be considered as reliable tool for the measure of the student's achievement in English.

Validity

Content validity was determined by showing that the behaviour demonstrated in testing constitute a representative sample of performance domain. The domain usually involves learned knowledge and skills. Thus, it is commonly used for achievement test (Wolf, 1982). The content validity is determined by comparing the items in a test with the content and objectives of a particular domain to see how well they match, as it is content of a particular direction. The test represents a fairly well-defined universe of content, content was made closely parallel to the tasks constituting the universe under study and performance on individual items was determined both with respect to the accuracy of the responses and the process used to solve the items.

REFERENCES

- Berk, R. A. (1980). Consumer guide to criterion reference test reliability. *Journal of Educational Measurement*, 17 (4), 323-349.
- Best, J. W., & Kahn, J. V. (1989). *Research in education* (6thed.). New Delhi: Prentice Hall.
- Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, 29 (7), 512-518.
- Dececco, J.P., & Crawford, W.R. (1974). *The psychology of learning and instruction*. Englewood Cliffs, NJ: Prentice Hall.

- Dowine, N.M., & Heath, R.W (1974). *Basic statistical methods* (4th ed.). New York: Harper & Row.
- Ebel, R.L. (1966). *Measuring educational achievement*. New Delhi: Prentice Hall.
- Emrick, J. (1971). An evaluation model for mastery testing. *Journal of Educational Measurement*, 8 (4), 321-326.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist* 18 (8), 519-521.
- Gronlund, N.E. (1977). *Constructing achievement tests*. New York: Prentice Hall.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw Hill.
- Hambleton, R. K., & Novick, M.R. (1973). Toward an integration of theory and method for criterion- referenced tests. *Journal of Educational Measurement*, 10 (3), 159-170.
- Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement*, 13 (4), 253-264.
- Ivens, S. (1970). *An investigation of item analysis, reliability, and validity in relation to criterion referenced tests*. Unpublished Doctoral Dissertation, Tallahassee: Florida, State University.
Retrieved May 3, 2019 from http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_197302_dziuban.pdf
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 40 (1), 17-19.
- Kriewall, T. E. (1972, April). *Aspects and applications of criterion- referenced tests*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois. Retrieved May 5, 2019 from <http://files.eric.ed.gov/fulltext/ED063333.pdf>
- Lindeman, R.H. (1967). *Educational measurement*. Bombay: D.B. Taraorevala Sons Private Ltd.
- Livingston, S. (1972). Criterion- referenced applications of classical test theory. *Journal of Educational Measurement*, 9 (1), 13-26.

- Mehar, R. (1997). *Role of advance organizer in learning and retention with respect to cognitive styles and learning types in Geography*. Unpublished Ph.D. Thesis, Chandigarh: Panjab University.
- Mouly, G. J. (1970). *The science of educational research*. New York: Van Nostrand Reinhold Company.
- Nunnally, J. C. (1982). Reliability of measurement. In H.E. Mitzel (Ed.), *Encyclopedia of educational research* (pp.1589-1601). New York: Free Press.
- Popham, W. J. (1975). *Educational evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Popham, W. J., & Husek, T. R. (1969). Implication of criterion referenced measurement. *Journal of Educational Measurement*, 6(1), 1-10.
- Shaycoft, M. F. (1979). *Handbook of criterion referenced testing*. New York: Garland STPM-Press.
- Stanley, J.C. (1964). *Measurement in todays schools* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Stanley, J.C., & Hopkins, K. D. (1972). *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Wolf, M. (1982). The word-retrieval process and reading in children and aphasics. In K. Nelson (Ed.), *Children's language*. New York: Gardner Press.