

Construction and Standardization of Achievement Test in English Language Acquisition Skills for Undergraduate Students of School of Open Learning (SOL), University of Delhi

Dr. Payal Mittal

Assistant Professor, Dept. of English, Swami Shraddhanand College, University of Delhi, New Delhi, India

Email Id: Payalmittal111@rediffmail.com

Dr. Parvesh Lata

Associate Professor & Head, Department of Education, School of Liberal Arts, G D Goenka University,
Gurugram, Harayana, India

Email. Id: Lataparvesh11@gmail.Com

Prof. Sarika Sharma

Professor, Department of Teacher Training and Non-Formal Education, Jamia Milia Islamia, New Delhi

Email Id: ssharma4@jmi.ac.in

Abstract

According to Elliott (2017), Achievement tests are seen to be crucial for evaluating future learning as it is the instruments used to evaluate how well students can apply the knowledge and skills taught and demonstrate acquisition. So it is very essential to standardize the achievement test through a pilot study which means the pre-testing or 'trying out' of a particular research tool. In the same context, the researcher has developed an achievement test for assessing the English language acquisition skills of undergraduate students of School of Open Learning (SOL), University of Delhi with different types of 66 questions. After pilot testing, item analysis was calculated with discrimination index and difficulty value. Only 45 questions were retained in the draft of achievement test. Validity and reliability of the achievement test was also calculated by using Spearman Brown Formula and the final draft of the achievement test was established.

Keywords: Pre Achievement Test, English Language Acquisition Skills, Item Analysis, Validity, Reliability

Introduction

Achievement is the degree of success and competence gained in a certain field with regard to academic work (Gupta & Lata, 2014). In other words, achievement is the competence of a person in relation to a domain of knowledge. Teaching-learning process is complete when its objectives are attained and measured in some measurable terms – qualitative or quantitatively. Achievement of learning objectives or goals may be seen in transformation of behaviours, gaining or improving a skill, increasing knowledge and understanding.

In a teaching-learning process, it is very crucial to assess and inform about the students' performance or achievements in relation to the standards, which constitutes the measure the objectives or aims. Various methods are used to assess or measure the individuals' knowledge, skill, attributes in the form of a written test, an observation, an oral test, an achievement test, assignment, standardized test, project, portfolio, etc. After that, the test score, both quantitative and qualitative acts as the evidence of reflection of the observations or achievement made during the test. Achievements test results are required to make many important decisions in the education system: to promote students to the next level, to check the efficiency of teaching-learning process, to check the learning gaps in students' knowledge and understanding of the specific domain, etc. Achievement test for English language acquisition needs to cover all the skills of the language which cannot be tested only through written exam. Anderson (1953) laid out some principles for constructing a good achievement test for English language skills:

1. Achievement should assess all the objectives of the teaching, if possible, assess each objective separately. In case of English language, where spoken language constitutes the main part of the learning a language but in India, it's the most ignored skills in most of the examination.
2. An achievement test should have questions from that materials taught in the class and must have more short questions than long ones. It should be long examination than short and should cover the questions ranging from easy to difficult.
3. An achievement test should be diagnostic. Since learning English mostly consists of developing skills and habits, the test should place a strong emphasis on actual mastery of the language skills. Exams with a 40% passing percentage and the majority of the

questions requiring in-depth knowledge of the language rather than proficiency in it should be discouraged.

4. An achievement test should make the good drilling exercises. It should requires the students to use the English language in a natural way, not in the translation way. Since meaning is dependent on the entire context in which a person is in, so the questions in an achievement test should be presented in context rather than as isolated words.

Objective of the study

The main objective of the current study is to construct and standardize an achievement test for English language acquisition skills for undergraduate students of BA (Prog.) and B.Com (Prog.), 2nd semester studying in School of Open Learning (SOL), University of Delhi.

Construction and Standardization of Achievement test

The researcher planned to assess the six English language acquisition skills – Reading, Listening, Speaking, Writing, Grammar, and Vocabulary. For attain this purpose, the researcher constructed an achievement test and standardize it so that it can measure all the skills of English language taught to the undergraduate students of School of Open Learning (SOL), University of Delhi. There are two categories of test: Teacher-made tests and Standardized test. Teacher-made tests are made by teachers to assess the student's progress in school. These tests are made for the specific content of the specific unit or specific course. Standardized test means the test that asks all examinees to answer the same questions or a set of questions in the same way and it is scored in a "standard" or consistent manner, allowing for comparison of the relative performance of individual students or groups of students. These standardized tests are widely used as they are the efficient way to evaluate the students' performance. The researcher could not able to relate to any appropriate standardized test so the there was a need to construct one such achievement test that needs to be standardized to some specific themes and can evaluate the learning of all the six skills mentioned above.

The researcher took the following steps to create and standardize an achievement test for assessing English language acquisition of undergraduate students of SOL:

1. Planning the test
2. Preparation of the test
3. Administration of the test (Pre try out and final try out)
4. Item analysis
5. Standardization of the test

1. Planning of the Test

For any process to be successful and productive, planning is the most important and vital step. The researcher considered these things for the planning of the achievement test: the six language skills to be assessed, the content area, the objectives, the themes, the type of questions, the time available, needs analysis of the students done by the researcher, etc. Further, the researcher plans to structure the achievement test on four domains:

1. Objectives of the test
2. Content of the test
3. Size and the types of questions
4. Preparation of the blueprint

1.1 Objectives of the test

The first and foremost step in constructing the achievement test is to define the objectives. Objectives are defined in the behavioral terms according to the Bloom's Taxonomy, focusing, only on knowledge, understanding, application, and skills. The weightage for each objective in the achievement test is given below in the Table 1.1 and also shown through pie chart in Fig. 1.1:

Table 1.1: Objective-wise weightage in the achievement Test

| Objective | Weightage |
|------------------|------------------|
| Knowledge | 27 |
| Understanding | 32 |
| Application | 31 |
| Skill | 10 |
| Total | 100 |

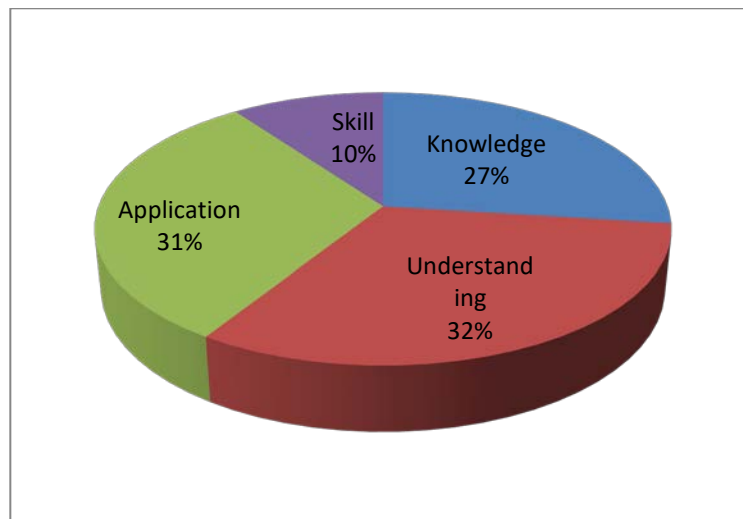


Fig. 1.1: Objective-wise weightage in the achievement Test

1.2 Content of the Test

For the content of achievement test, the researcher focusing on measuring the six English language skills: Reading, listening, speaking, writing, grammar, and vocabulary. Many readily available standardized tests only assess reading, writing, grammar, and a little amount of vocabulary, leaving out the other two crucial language skills of speaking and listening. As a result, they are not very thorough in assessing skills. For this reason, the researcher forced to create and standardize an achievement test. According to the researcher's need analysis, it is revealed that students mostly want exposure and practice in speaking and listening. Skill-wise weightage is shown in the Table 1.2 and also through pie chart in the Fig. 1.2:

Table 1.2: Skill-wise weightage in the achievement Test

| Skill | Weightage |
|------------|-----------|
| Reading | 15 |
| Listening | 15 |
| Writing | 16 |
| Speaking | 9 |
| Grammar | 30 |
| Vocabulary | 15 |
| Total | 100 |

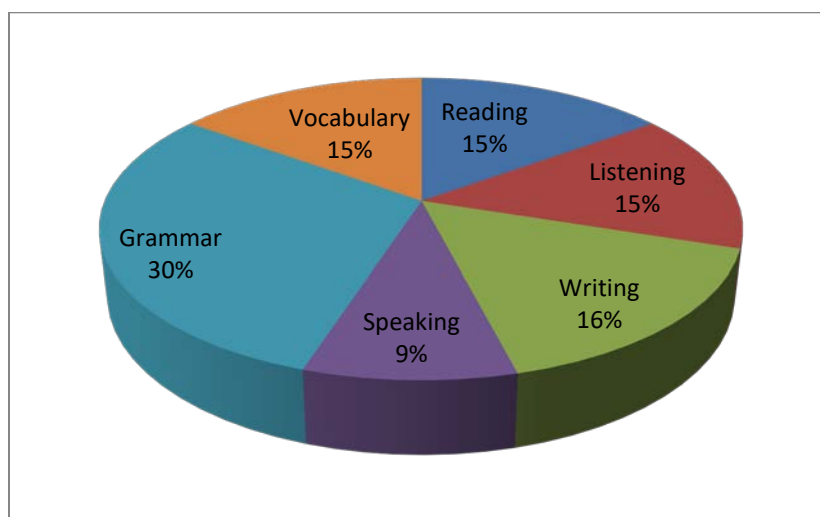


Fig. 1.2: Skill-wise weightage in the achievement Test

1.3 Size and Types of Questions

The next step is to select the number of test items in the achievement test and the types of questions for the test items. The researcher consulted many English language experts for the size and types of questions included in the achievement test. Initially, 66 questions were put in the first draft of the achievement test. To measure all the skills of English language, it has different types of questions: Multiple choice questions, Fill in the blanks, and Long questions. The question type weightage is given in the table 1.3:

Table 1.3: Question type-Weightage in the Achievement Test

| Types of Questions | Weightage |
|--------------------|-----------|
| Multiple choice | 56 |
| Fill in the blanks | 7 |
| Long question | 3 |
| Total | 66 |

1.4 Preparation of the Blue print

A blueprint for an achievement test is an organized document that directs the construction of a test. It ensures that the test accurately evaluates the desired learning outcomes by outlining the test's goals, subject areas, and question forms. In the blueprint, the researcher allots the number of

questions and marks of those questions according to the time given. The researcher presents its final decision in the form of blueprint given below in the Table 1.4:

Table 1.4: Blueprint for Pre-Tryout Achievement Test

| Objectives | Knowledge | | | Understanding | | Application | | | Skills | | | Total |
|-------------------|-----------|------|--|---------------|------|-------------|-------|------|--------|----|------|----------------|
| Forms of Question | MCQ | FB | | MCQ | FB | MCQ | FB | LA | MCQ | FB | LA | |
| Topic | | | | | | | | | | | | |
| Reading | 6(6) | | | 6(6) | | 3(3) | | | | | | 15(15) |
| Writing | 3(3) | | | 4(1) | | | 4(1) | | | | 5(1) | 16(6) |
| Vocabulary | 4(4) | | | 5(5) | | | 6(1) | | | | | 15(10) |
| Listening | 4(4) | | | 8(8) | | | 2(2) | | 1(1) | | | 15(15) |
| Speaking | | | | | 3(1) | | 2(1) | | | | 4(1) | 9(3) |
| Grammar | 2(1) | 8(1) | | 6(6) | | 8(8) | | 6(1) | | | | 30(17) |
| Total | 19(18) | 8(1) | | 29(26) | 3(1) | 11(11) | 14(5) | 6(1) | 1(1) | | 9(2) | 100(66) |

MCQ: Multiple Choice Questions

FB: Fill in the Blanks

LA: Long Answer Question

The numbers in brackets are the number of questions.

The numbers outside the brackets are the marks assigned to these questions

2. Preparation of the test

After incorporating the suggestions of language experts, the initial draft of the achievement test was constructed and finalized. In this draft, there are total 66 questions of three types: multiple choice, fill in the blanks, and long questions. These questions were constructed corresponding to the 5 units based on different themes like Travel, Entertainment, Health and Food, Technology, and Workplace. These questions aim towards to evaluate the efficiency of six skills of English language: reading, listening, writing, speaking, grammar, and vocabulary. Respondents are required to check the appropriate response out of four options when answering multiple-choice

questions. Respondents must supply a word or phrase to finish the statement in fill-in-the-blanks. They have to write a coherent paragraph in writing skills and speak in response to an inquiry in the speaking skill section.

Simple, appropriate, and easy-to-understand instructions were prepared for the directions for attempting the test items for the students. In the beginning of each section, clear and detailed instructions were given for students' understanding. Along with it, a detailed and concise instructions were also provided for how to administrate the achievement test. At the end, scoring key was prepared for the scoring process.

3. Administration of the test

Now, the achievement test for English language acquisition skills ready to be administered with 66 questions and scoring key. The achievement test was administered in two phases: Preliminary tryout and final tryout.

3.1 Preliminary Tryout

The preliminary tryout of an achievement test is defined as a small-scale administration of an achievement test to a group of students who are representative of the target population but not the final test group. Before the achievement test is administered, it is important to assess its validity, clarity, level of difficulty, unclear sentences, format, etc. through the preliminary tryout. For this, the preliminary draft of achievement test was administered on 20 students.

All the problems and errors were noted and further the 10 questions were reframed for ease of the students.

3.2 Final Tryout

In the final tryout, 45 students of BA (Prog.) of 2nd semester were selected who had taken the Modern Indian Language course "Language through Literature" in the first semester. The researcher's original goal was to enroll 90 students, but only 45 were convinced because the test was a little lengthy and the students didn't have much free time. Every student received the test and they were expected to mark or write their responses in the space provided in the achievement test. The test was of two hours. There were three types of questions: In multiple choice question,

the students had to tick the right answer out of four options; in fill in the blanks type question, they were supposed to complete the sentence by a word or a phrase; in long question, they were expected to write a coherence paragraph and answer some questions orally for the speaking skill section.

3.3 Scoring

When the students completed the achievement test, the answer sheets were collected. Scores were given on the basis of scoring key. The answers to multiple choice question and fill in the blanks were given one mark for the correct answer and zero mark for the incorrect and unattempt questions. Long questions were scored according to rubric made beforehand. This is how all the students' answers sheets were scored.

4. Item analysis

Item analysis is the process of assessing each test question's (items') quality by looking at how students answer it. It assists in determining whether an item is too easy, too difficult, or badly phrased, as well as how well it distinguishes between students with varying skill levels. Item analysis aims at to give more valid and reliable test by identifying issues with individual test items, give insight about the improvement in item's instruction according to students' strengths and weaknesses, assesses whether the item should be removed or revised. So, item analysis helps to improve the quality of achievement test by identifying the weak item.

In other words, item analysis is the statistical method used to choose and reject test items based on their discrimination index and difficulty value. Microsoft Excel was used by the researcher to conduct the item analysis, which simplifies computation and analysis.

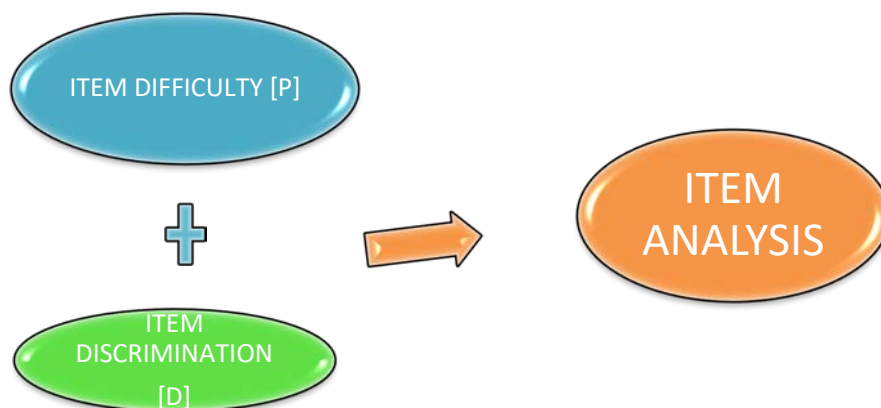


Fig. 1.3: Item Analysis on the basis of Item Difficulty and Item Discrimination

The following steps were taken to do the item analysis:

- (i) Each respondent's average score was determined, and the respondents were ranked descending order using their average scores.
- (ii) Next, the formula—27% of all respondents—was used to determine the "upper group" and the "lower group." Thus, there were 45 responders in all (27% of 45 = 12.1 and

The average scores have previously been sorted in descending order. The first 12 respondents' total scores for each item or question are designated as the upper group (UG), and the final 12 respondents' scores are designated as the lower group (LG). Of the answer papers, the middle 46% were set aside.

- (iii) Determine the difficulty level: After created the "upper group" and "lower group," each respondent's scores were recorded and tabulated. The thing will be easier if the overall number is higher. The percentage of students who properly answered the item was used to determine its difficulty. According to J P Guilford, "Difficulty value of an item is defined as the proportion or percentage of the examinees who have answered the items correctly." The percentage was converted into proportions. The average of the proportions of correct responses on each item in the two end groups was taken to be an estimate of the difficulty value of that particular item.

Formula for computing the difficulty value 'dv' of each item was:

$$dv = P_U + P_L / 2 ,$$

where, dv = difficulty value of the items.

P_U = proportions of correct responses to the items from the upper group.

P_L = proportions of correct responses to the items from the lower group.

- (iv) Internal Consistency Discrimination Index (rb): Internal consistency discrimination index (rb) of an item is the relationship between the total scores obtained from a test and item scores. This was discovered by examining the J.C. Flanagan's abac's bi-serial coefficient of correlation between item and overall score. Flanagan's abac was created to be used when the sample was limited to the top and bottom 27% of the distribution of total scores, and the middle 46% of examinees were removed. The ordinate displayed the percentage of examinees in the upper criterion group who passed the item, while the abscissa displayed the comparable percentage from the lower criterion group. The intersection of perpendiculars yielded the value of the coefficient rb. Items for the final draft were chosen once each item's difficulty levels and internal consistency discrimination indices were established, as previously mentioned.

$$DV = (UG + LG)/(NH + NL)$$

Whereas, UG = Average score of that item

LG = Average score of that item

NH = Number of respondents in highest group

NL = Number of respondents in lower group

4.1 Item Selection for Final Draft

The discrimination index and difficulty value of each item were taken into account to choose the final items. Item difficulty, according to Karadag & Sahin (2016), is the proportion of students who answered an item correctly; it varies from 0.0 to 1.0. The capacity to differentiate between students with high and poor scores is reflected by an item's discrimination index, which increases in difficulty as the item's difficulty value approaches zero.

When the item's value is nearer 1, it is considered good because it can distinguish between pupils who score highly and those who score poorly. Each item's discriminating and difficulty indices are examined to give teachers insight into what students have learnt and to assist them pinpoint and address any issues. In other words, it contributes to improving the validity and reliability of the tests by revealing whether or not the items are operating correctly.

The Table 1.5 shows the distribution of the items finally selected according to the 'dv' and 'rb':

Table 1.5: Distribution of the Items Selected on the basis of the 'dv' and 'rb':

| Test Items | UG | LG | DV | Interpretation | Disc. Index | Interpre. | decision |
|------------|----|----|-------------|----------------|-------------|-----------------|----------|
| 1 | 6 | 1 | 0.155555556 | VERY DIFFICULT | 0.4166667 | VERY GOOD | ACCEPT |
| 2 | 6 | 6 | 0.266666667 | AVERAGE | 0 | POOR ITEM | REJECT |
| 3 | 10 | 2 | 0.266666667 | AVERAGE | 0.6666667 | VERY GOOD | RETAIN |
| 4 | 5 | 4 | 0.2 | VERY DIFFICULT | 0.0833333 | POOR ITEM | REJECT |
| 5 | 8 | 10 | 0.4 | AVERAGE | -0.166667 | POOR ITEM | REJECT |
| 6 | 9 | 6 | 0.333333333 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 7 | 12 | 9 | 0.466666667 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 8 | 12 | 7 | 0.422222222 | AVERAGE | 0.4166667 | VERY GOOD | RETAIN |
| 9 | 12 | 2 | 0.311111111 | AVERAGE | 0.8333333 | VERY GOOD | RETAIN |
| 10 | 10 | 7 | 0.377777778 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 11 | 3 | 3 | 0.133333333 | VERY DIFFICULT | 0 | POOR ITEM | REJECT |
| 12 | 12 | 5 | 0.377777778 | AVERAGE | 0.5833333 | VERY GOOD | RETAIN |
| 13 | 2 | 2 | 0.088888889 | VERY DIFFICULT | 0 | POOR ITEM | REJECT |
| 14 | 11 | 6 | 0.377777778 | AVERAGE | 0.4166667 | VERY GOOD | RETAIN |
| 15 | 10 | 7 | 0.377777778 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 16 | 12 | 12 | 0.533333333 | AVERAGE | 0 | POOR ITEM | REJECT |
| 17 | 12 | 11 | 0.511111111 | AVERAGE | 0.0833333 | POOR ITEM | REJECT |
| 18 | 11 | 9 | 0.444444444 | AVERAGE | 0.1666667 | POOR ITEM | REJECT |
| 19 | 10 | 7 | 0.377777778 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 20 | 11 | 6 | 0.377777778 | AVERAGE | 0.4166667 | VERY GOOD | RETAIN |
| 21 | 10 | 4 | 0.311111111 | AVERAGE | 0.5 | VERY GOOD | RETAIN |
| 22 | 8 | 6 | 0.311111111 | AVERAGE | 0.1666667 | POOR ITEM | REJECT |
| 23 | 11 | 3 | 0.311111111 | AVERAGE | 0.6666667 | VERY GOOD | RETAIN |
| 24 | 11 | 10 | 0.466666667 | AVERAGE | 0.0833333 | POOR ITEM | REJECT |
| 25 | 7 | 4 | 0.244444444 | VERY DIFFICULT | 0.25 | MARGINAL ITEM | RETAIN |
| 26 | 11 | 8 | 0.422222222 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 27 | 3 | 2 | 0.111111111 | VERY DIFFICULT | 0.0833333 | POOR ITEM | REJECT |
| 28 | 12 | 8 | 0.444444444 | AVERAGE | 0.3333333 | REASONABLY GOOD | RETAIN |

| | | | | | | | |
|----|----|----|-------------|----------------|-----------|--------------------|--------|
| 29 | 12 | 9 | 0.466666667 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 30 | 6 | 5 | 0.244444444 | VERY DIFFICULT | 0.0833333 | POOR ITEM | REJECT |
| 31 | 31 | 20 | 1.133333333 | VERY EASY | 0.9166667 | VERY GOOD | RETAIN |
| 32 | 12 | 5 | 0.377777778 | AVERAGE | 0.5833333 | VERY GOOD | RETAIN |
| 33 | 12 | 8 | 0.444444444 | AVERAGE | 0.3333333 | REASONABLY GOOD | RETAIN |
| 34 | 6 | 2 | 0.177777778 | VERY DIFFICULT | 0.3333333 | REASONABLY GOOD | RETAIN |
| 35 | 10 | 6 | 0.355555556 | AVERAGE | 0.3333333 | REASONABLY GOOD | RETAIN |
| 36 | 12 | 11 | 0.511111111 | AVERAGE | 0.0833333 | POOR ITEM | REJECT |
| 37 | 12 | 9 | 0.466666667 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 38 | 12 | 6 | 0.4 | AVERAGE | 0.5 | VERY GOOD | RETAIN |
| 39 | 9 | 5 | 0.311111111 | AVERAGE | 0.3333333 | REASONABLY GOOD | RETAIN |
| 40 | 9 | 10 | 0.422222222 | AVERAGE | -0.083333 | POOR ITEM | REJECT |
| 41 | 36 | 8 | 0.977777778 | VERY EASY | 2.3333333 | VERY GOOD | RETAIN |
| 42 | 67 | 17 | 1.866666667 | VERY EASY | 4.1666667 | VERY GOOD | RETAIN |
| 43 | 12 | 8 | 0.444444444 | AVERAGE | 0.3333333 | REASONABLY GOOD | RETAIN |
| 44 | 9 | 1 | 0.222222222 | VERY DIFFICULT | 0.6666667 | VERY GOOD | RETAIN |
| 45 | 4 | 3 | 0.155555556 | VERY DIFFICULT | 0.0833333 | POOR ITEM | REJECT |
| 46 | 3 | 4 | 0.155555556 | VERY DIFFICULT | -0.083333 | POOR ITEM | REJECT |
| 47 | 11 | 4 | 0.333333333 | AVERAGE | 0.5833333 | VERY GOOD | RETAIN |
| 48 | 20 | 9 | 0.644444444 | AVERAGE | 0.9166667 | VERY GOOD | RETAIN |
| 49 | 12 | 5 | 0.377777778 | AVERAGE | 0.5833333 | VERY GOOD | RETAIN |
| 50 | 6 | 4 | 0.222222222 | VERY DIFFICULT | 0.1666667 | POOR ITEM | REJECT |
| 51 | 10 | 7 | 0.377777778 | AVERAGE | 0.25 | MARGINAL ITEM | RETAIN |
| 52 | 9 | 3 | 0.266666667 | AVERAGE | 0.5 | VERY GOOD | RETAIN |
| 53 | 8 | 1 | 0.2 | VERY DIFFICULT | 0.5833333 | VERY GOOD | RETAIN |
| 54 | 10 | 8 | 0.4 | AVERAGE | 0.1666667 | POOR ITEM | REJECT |
| 55 | 9 | 4 | 0.288888889 | AVERAGE | 0.4166667 | VERY GOOD | RETAIN |
| 56 | 11 | 5 | 0.355555556 | AVERFAGE | 0.5 | VERY GOOD | RETAIN |
| 57 | 5 | 8 | 0.288888889 | AVERAGE | -0.25 | POOR ITEM | REJECT |
| 58 | 52 | 8 | 1.333333333 | VERY EASY | 3.6666667 | VERY GOOD | ACCEPT |
| 59 | 32 | 17 | 1.088888889 | VERY EASY | 1.25 | VERY GOOD | ACCEPT |
| 60 | 6 | 1 | 0.155555556 | VERY DIFFICULT | 0.4166667 | VERY GOOD | ACCEPT |
| 61 | 7 | 5 | 0.266666667 | AVERAGE | 0.1666667 | POOR ITEM | REJECT |
| 62 | 5 | 2 | 0.155555556 | VERY DIFFICULT | 0.25 | MARGINAL ITEM | ACCEPT |
| 63 | 23 | 7 | 0.666666667 | AVERAGE | 1.3333333 | VERY GOOD | RETAIN |

| | | | | | | | |
|----|----|----|-------------|-----------|-----------|-----------|--------|
| 64 | 46 | 22 | 1.511111111 | VERY EASY | 2 | VERY GOOD | ACCEPT |
| 65 | 34 | 12 | 1.022222222 | VERY EASY | 1.8333333 | VERY GOOD | ACCEPT |
| 66 | 17 | 0 | 0.377777778 | AVERAGE | 1.4166667 | VERY GOOD | RETAIN |

Ebel's (1979) criteria and guidelines for classifying discriminating indices are applied in order to analyze the test items in this achievement test.

Table 1.6 Ebel's Guidelines (Discriminating Powers)

| Discriminating powers | Description |
|------------------------------|---|
| .40 and above | The item is functioning quite satisfactorily |
| Between 0.30-0.39 | Little or no revision is required |
| Between 0.20-0.29 | The item is marginal and needs revision |
| <.19 | The item should be eliminated or completely revised |

Based on Ebel's guidelines mentioned in the above Table 1.6, the 66 test items were categorized in the Table 1.7 and 1.8 given below:

Table 1.7: Distribution of Discrimination Powers of all the Items of Achievement Test

| Discriminating Power | Frequency | Item Numbers | Remarks |
|-----------------------------|------------------|--|-------------------|
| .40 and above | 29 | 1, 3, 8, 9, 12, 14, 20, 21, 23, 31, 32, 38, 41, 42, 44, 47, 48, 49, 52, 53, 55, 56, 58, 59, 60, 63, 64, 65, 66 | Very Good Items |
| .30-.39 | 06 | 28, 33, 34, 35, 39, 43 | Reasonably Good |
| .20-.29 | 11 | 6, 7, 10, 15, 19, 25, 26, 29, 37, 51, 62 | Needs Improvement |
| <.19 | 20 | 2, 4, 5, 11, 13, 16, 17, 18, 22, 24, 27, 30, 36, 40, 45, 46, 50, 54, 57, 61 | Very Poor Item |
| Total | 66 | | |

**Table 1.8: Bi-Variate Scatter Diagram for Achievement Test in English
Language Acquisition Skills between Difficulty Value (dv) and Discrimination
Index (rb)**

| Rbdv | .00-.09 | .10-.19 | .20-.29 | .30-.39 | .40-.49 | .50-.59 | .60-.69 | .70-.79 | .80-.89 | .90-.99 |
|---------|----------|-------------|-----------|-------------------|------------|----------|---------|---------|----------------|---------|
| .00-.09 | 13,45,57 | 11,27,46,54 | 2,4,30,61 | | 5,24,40 | 16,17,36 | | | | |
| .10-.19 | | | 50 | 22 | 18 | | | | | |
| .20-.29 | | 62 | 25 | 6,10,15,19,51 | 7,26,29,37 | | | | | 41 |
| .30-.39 | | 34 | | 35,39 | 28,33,43 | | | | | |
| .40-.49 | | 1,60 | 55 | 14,20 | 8 | | | | | |
| .50-.59 | | | 52,53 | 12,21,32,47,49,56 | 38 | | | | | |
| .60-.69 | | | 3,44 | 23 | | | | | | |
| .70-.79 | | | | | | | | | | |
| .80-.89 | | | | 9 | | | | | | |
| .90-.99 | | 58 | | 66 | | | 48,63 | | 31,42,59,64,65 | |

Tables 1.7 and 1.8 show that 17 items have discriminating powers between .20 and .39, whereas 29 items have discriminating powers of .40 or higher; as a result, they were chosen for the final draft. These were updated with regard to item difficulty level, content complexity, and language clarity. The wording was made clearer and simpler to understand. Uncertain things were explained. Ultimately, these elements were incorporated into the final copy once more detailed instructions were given for trying them. The remaining 20 items were excluded because their discriminating power was less than .19.

5. Reliability of the Achievement Test

The researcher used the split-half methodology to determine the reliability. By dividing the test items into odd and even-numbered items, the things have been divided into two tests that are equal in content and complexity. The reliability coefficient, which the researcher assessed to be 0.833, indicates that true-score variation accounts for 83.3% of test score variance while error variance accounts for just 16.7%. According to the general rules for assessing test reliability listed below in Tables 1.9 and 1.10, we can therefore conclude that the accomplishment test is reasonably reliable.

Table 1.9: Reliability Test for Achievement Test

| Reliability Statistics | | | |
|--------------------------------|------------------|------------|-----------------|
| Cronbach's Alpha | Part 1 | Value | .772 |
| | | N of Items | 23 ^a |
| | Part 2 | Value | .855 |
| | | N of Items | 23 ^b |
| | Total N of Items | | 46 |
| Correlation Between Forms | | | .714 |
| Spearman-Brown Coefficient | Equal Length | | .833 |
| | Unequal Length | | .833 |
| Guttman Split-Half Coefficient | | | .694 |

a. The items are: Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7, Q_8, Q_9, Q_10, Q_11, Q_12, Q_13, Q_14, Q_15, Q_16, Q_17, Q_18, Q_19, Q_20, Q_21, Q_22, Q_23.

b. The items are: Q_24, Q_25, Q_26, Q_27, Q_28, Q_29, Q_30, Q_31, Q_32, Q_33, Q_34, Q_35, Q_36, Q_37, Q_38, Q_39, Q_40, Q_41, Q_42, Q_43, Q_44, Q_45, Q_46.

Table 1.10: General Guidelines for Interpreting Reliability Coefficients.

| S.No | Reliability coefficient value | Interpretation |
|------|-------------------------------|--------------------------------|
| 1. | 0.90 and above | Excellent |
| 2. | 0.80 – 0.89 | Good |
| 3. | 0.70 – 0.79 | Adequate |
| 4. | Below 0.70 | May have limited applicability |

6. Validity of the Achievement Test

The use of comprehensive statistical methods for test validation was constrained by the nature of the test questions and the goal of the current inquiry. The Content Validity criterion was used to determine validity. The suitability of sampling a given universe of content is the focus of content validity. To determine the validity of the material, a panel of three experts in test items and five subject matter experts, was given a list of outcomes along with the test items. The panel was asked to identify which learning result was associated with which test item. In 95% of cases, the experts agreed with the researcher about the assignment of test items to objectives. The percentage was taken as evidence of Content Validity.

7. Final Draft of the Achievement Test

The achievement test's items were finalized and rearranged accordingly. The instructions on how to take the test were given on the cover page. Along with the achievement test, the scoring key was prepared, which was based on the test's final draft. There are 31 test items on the final achievement test which needs to be completed in two-hour time. The blueprint for the final

achievement test is given in Table 1.11, in which the test items for each skill were arranged and presented according to the different cognitive levels.

Table 1.11: Blueprint for the Final Achievement Test

| Objectives | Knowledge | | | Understanding | | | Application | | | Skills | | | Total |
|-------------------|-----------|------|----|---------------|------|----|-------------|----|------|--------|----|------|--------|
| Forms of Question | QB | FB | LQ | QB | FB | LQ | QB | FB | LQ | QB | FB | LQ | |
| Topic | | | | | | | | | | | | | |
| Reading | 2(2) | | | 3(3) | | | 2(2) | | | | | | |
| Writing | 1(1) | | | 1(1) | | | | | | | | 5(1) | 7(3) |
| Vocabulary | | | | 5(5) | | | | | | | | | 5(5) |
| Listening | 4(4) | | | 2(2) | | | | | | | | | 6(6) |
| Speaking | | | | | 3(1) | | | | | | | 4(1) | 7(2) |
| Grammar | 2(1) | 8(1) | | 1(1) | | | 4(4) | | 6(1) | | | | 21(8) |
| Total | 9(8) | 8(1) | | 12(12) | 3(1) | | 6(6) | | 6(1) | | | 9(2) | 53(31) |

LA: Long Answer Question

MCQ: Multiple Choice Question

FB: Fill in the Blanks

The numbers in brackets are the number of questions.

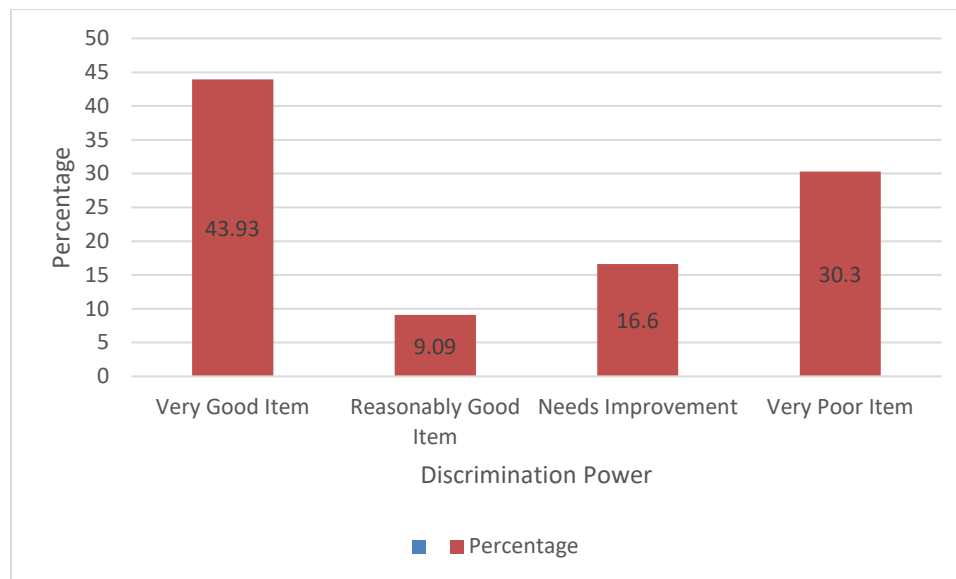
The numbers outside the brackets are the marks assigned to these questions

Discussion

Tables 1.8 and 1.9 demonstrate that 29 items (43.93%) have discriminating power. The final draft had 29 items with Discriminating Power ranging from .20 to .39, which were chosen because they were 40 or above. 6 items or (9.09%) were added to the final document after being edited and changed. The remaining 20 items were excluded because their discriminating power was less than .19 (30.3%). 11 test items or (16.6%) with discriminating powers ranging from .20 to .39 were altered based on the items' degree of difficulty, content complexity, and linguistic clarity. Fig. 1.4 provides a bar diagram that displays every item based on discrimination power. The wording was

simplified to make it easier to understand. Items that were vague were clarified. Clearer instructions were provided on how to attempt the questions.

Fig. 1.4: Bar Diagram shows Item based on Discrimination Power



Conclusion

The quality of an achievement test is determined by the quality of test items included in it. For this purpose, item analysis is used. Item analysis is used to improve the quality of an achievement by removing the unclear or misleading test items and identifying the areas or test items where the more clarity is required. So, item analysis is an important phase in developing a test or instrument. The findings of this paper is very relevance for the test developers, teachers, and student teachers. This achievement test includes all the language skills in English to make it more dynamic and holistic.

References

Anderson, D. F. (1953). Tests of Achievement in English Language. *ELT Journal*, VII (2), 37-69.
10.1093/elt/vii.2.37.

Gupta, M., & Lata, P. (2017). Effectiveness of IT-enabled Instructional package (ITEIP) on science achievement of X class students in relation to their gender. *British Journal of Multidisciplinary and Advanced Studies*, 1(1), 30-44.

Kocdar, S., Karadag, N., & Sahin, M. D. (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context. *Turkish Online Journal of Educational Technology-TOJET*, 15(4), 16-24.

Sharma, H L and Poonam (2017). Construction and standardization of an achievement test in English grammar. *International Journal of Advanced Educational Research*, 2(5), 230-235. ISSN: 2455-6157.

https://www.researchgate.net/publication/342918149_ASSESSMENT_AND_EVALUATION_IN_EDUCATION

Stephen N. Elliott (2017), Achievement Tests, Reference Module in Neuroscience and Biobehavioral Psychology, Elsevier, ISBN 9780128093245, <https://doi.org/10.1016/B978-0-12-809324-5.05457-2>.

(<https://www.sciencedirect.com/science/article/pii/B9780128093245054572>)