



---

## Development And Standardisation Of English Language Proficiency Test For 8<sup>th</sup> Grade Students

**Hemant Lata Sharma** Professor, Department of Education, MDU, Rohtak, Haryana, India. Email: [piscean\\_hemant@yahoo.com](mailto:piscean_hemant@yahoo.com)

**\*\*Kiran Rani\*\*** (Corresponding Author) Assistant Professor, School of Education, Central University of Haryana, Mahendergarh, Haryana, India. Email: [kiranshayna@gmail.com](mailto:kiranshayna@gmail.com)

---

### ABSTRACT

In a rapidly globalizing world, the English language has established itself as a lingua-franca internationally among the speakers of different languages as well as in a multilingual country like India. Thus, in this scenario, learners of English as a Second Language (ESL) need to possess not only the knowledge of the English language, but also the ability to use it for practical purposes. But it is usually seen that in Indian classrooms, barring a few exceptions as in the case of some well-equipped, resource-rich schools, ESL teaching-learning and assessment focuses more on the development and assessment of Reading and Writing skills with hardly any scope for the enhancement and assessment of the Listening and Speaking skills. This is also evident from the written-test (achievement test) format for the evaluation of language-subjects rather than proficiency-test format which is more suitable for a comprehensive assessment in language. But the importance and relevance of proficiency tests, is emphasized by the National Focus Group Position Paper on Teaching of English-NCERT (2006) too, wherein it is stated that “Language evaluation need not be tied to ‘achievement’ with respect to particular syllabi, but must be reoriented to the measurement of language proficiency. National benchmarks for language proficiency need to be evolved”. Against this backdrop, a need was felt by the researcher to develop and standardize a proficiency test for English language skills for 8<sup>th</sup> grade students and this paper discusses the process of development and standardization of English Language Proficiency Test (ELPT).

**Keywords:** Development and Standardisation, English Language Proficiency Test, Language Evaluation, Language Skills’ Assessment.

### INTRODUCTION

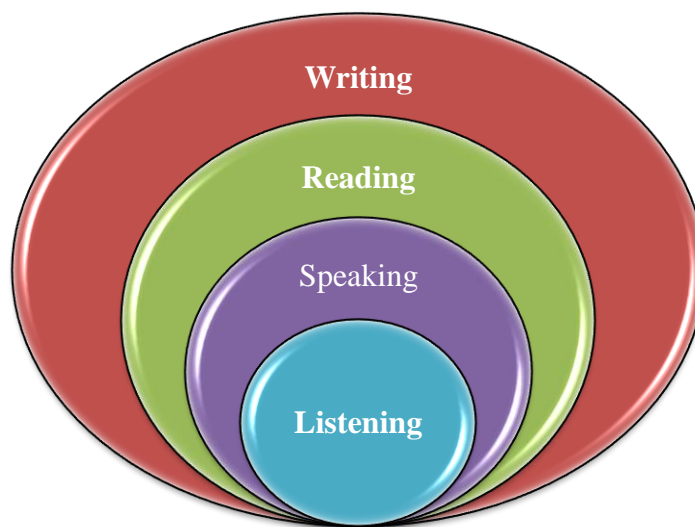
In a rapidly globalizing world, the English language has established itself as a lingua-franca internationally among the speakers of different languages as well as in a

multilingual country like India. Thus, in this scenario, learners of English as a Second Language (ESL) need to possess not only the knowledge of the English language, but also the ability to use it for practical purposes. But it is usually seen that in Indian classrooms, barring a few exceptions as in the case of some well-equipped, resource-rich schools, ESL teaching-learning and assessment focuses more on the development and assessment of Reading and Writing skills with hardly any scope for the enhancement and assessment of the Listening and Speaking skills. This is also evident from the written-test (achievement test) format for the evaluation of language-subjects rather than proficiency-test format which is more suitable for a comprehensive assessment in language. At present, the assessment pattern of languages is largely the same as the evaluation of knowledge-based subjects such as History, Geography etc. To understand this distinction between Achievement Test and Proficiency Test, we can say that 'Achievement' is assessed in terms of how well the students know what they are expected to know whereas 'proficiency test' measures what an individual can do with what he or she knows. This can be well understood through the example of a driver's test - the paper-and-pencil questions on the knowledge related to driving and traffic rules represent the Achievement Test, while a Proficiency Test determines how well the person can drive a car. This significant distinction between these two types of tests, as also the importance and relevance of proficiency tests, is recognized by the National Focus Group Position Paper on Teaching of English-NCERT (2006) too, wherein it is stated that "Language evaluation need not be tied to 'achievement' with respect to particular syllabi, but must be reoriented to the measurement of language proficiency..... National benchmarks for language proficiency need to be evolved preliminary to designing a set of optional English Language Tests that will balance curricular freedom with the standardisation of evaluation that certification requires, and serve to counter the current problem of English (along with mathematics) being a principal reason for failure at Class X. A student may be allowed to 'pass without English' if an alternative route for English certification (and therefore instruction) can be provided outside the regular school curriculum".

Based on what the above-mentioned Position Paper has stated, it can be argued that for the advancement of learners in terms of accumulation of more building blocks of language, the teaching-learning and assessment process should encourage the learners to respond in unexpected and real-world situations. Gradually, tests should be based on level-appropriate topics that are unplanned and potentially unpracticed, and assess the proficiency rather than achievement of a student. Against this backdrop, a need was felt by the researcher to develop and standardize a proficiency test for English language skills for 8<sup>th</sup> grade students and in the following discussion, the process of development and standardization of English Language Proficiency Test (ELPT) has been described in detail.

## **CONSTRUCTION AND STANDARDISATION OF ENGLISH LANGUAGE PROFICIENCY TEST (ELPT): REQUIREMENT, FACTORS TO BE CONSIDERED AND SUMMARY OF PROCEDURES ADOPTED**

An important takeaway from the National Focus Group Position Paper on Teaching of English-NCERT (2006) is the suggestion that “National benchmarks for language proficiency need to be evolved preliminary to designing a set of optional English Language Tests that will balance curricular freedom with the standardisation of evaluation....”. This makes it evidently clear that there is a need to develop and standardize English language proficiency tests for different levels of students. Thus, considering the gap between the requirement and availability, the researcher took up the endeavour to develop and standardize the English Language Proficiency Test (ELPT) for 8<sup>th</sup> grade students in order to assess their English language skills i.e. Listening, Speaking, Reading and Writing.



**Fig-4.1 Skills covered in ELPT**

Construction of a proficiency test is, however, a more tedious task as compared to constructing an achievement test. This is so, because, a proper, reliable, efficiently measurable testing of the skills would require technical paraphernalia which consists of audio playing and recording devices and an arrangement in which individual students' skills can be tested rather than the whole class collectively. It is obvious that the listening and speaking skills cannot be tested in a group though reading and writing can be. Further, the Reading skill has two components - Reading Comprehension and Reading Aloud Ability – and the latter has also to be tested individually though the former can be tested in a group. Therefore, a way had to be devised for administering the test even as all these considerations are kept in mind. Further, the format of assessment for each and every skill needed to be decided beforehand. The testing of listening and reading skills could be done by using the multiple-choice question-answer format but the testing of speaking and writing skills could be done only through seeking descriptive answers whereas maintaining equal standards for scoring descriptive answers is not easy. Therefore, in order to guide the scoring of every skill, initially, 10-point grading scales were constructed for each of the skills i.e. Listening, Speaking, Reading Comprehension, Reading Aloud Ability and Writing. But after obtaining the opinion of experts, these scales were modified and 6-point grading scales were constructed.

The first draft of the proficiency test was scored by following the six-point grading scale. The next important task was the selection of items that cover the various components of a skill. The proficiency test was pilot-tested and the reliability and validity were rigorously assessed at each stage as it is crucial to check whether the test will work or not. The objective, in this way, is to put the investigation to test, initially on a small scale, in order to deal with all the plausible issues that may cause disappointment later on. After thorough pilot-testing, ultimately 28-questions made it to the final version of the English Language Proficiency Test to assess English Language Skills.

### **POPULATION AND SAMPLE**

The students of elementary classes of government schools of the Mahendergarh district of Haryana constituted the population for the pilot study. A sample of 120 students studying in class 8<sup>th</sup> was randomly selected and administered the first draft of English Language Proficiency test for conducting Item-Analysis. Again, a sample of 120 students studying in class 8<sup>th</sup> of government schools of the Mahendergarh district of Haryana were administered the final form of the ELPT. It constituted the main developmental sample for standardization of the English Language Proficiency test.

### **STAGES OF CONSTRUCTION OF THE ENGLISH LANGUAGE PROFICIENCY TEST (ELPT)**

As stated earlier, the construction of the English Language Proficiency Test (ELPT) required the researcher to keep multiple considerations in mind. To meet the needs of the study and obtain reliable and valid results through good test, the researcher planned and executed the following stages to construct and to standardize the test.

- Preparing a 6-point grading scale for grading/scoring of English Language Proficiency of all the basic language skills i.e. Listening, Speaking, Reading and Writing.
- Preparing a pool of questions to assess English Language skills.
- Seeking experts' opinion on Questions and Grading Scale.
- Item Analysis of the test
- Seeking experts' opinion on the final format of ELPT
- Piloting of the test.

The final version of the English Language Proficiency Test was obtained as a result of meticulous work, paying attention to each and every detail of the purpose, process and product. The following objectives were framed to standardize the English Language Proficiency Test (ELPT) for 8<sup>th</sup> Grade students.

### **Objectives**

- I.** To construct a Skill-wise Grading Scale for grading/scoring of ELPT.
- II.** To construct the Questions of English Language Proficiency Test (ELPT) for 8<sup>th</sup> Grade students.
- III.** To find out the test scores of 8<sup>th</sup> Grade students.

**IV.** To find out the discriminating power of questions by performing the Item-analysis to establish the validity of all the drafts of the English Language Proficiency Test (ELPT).

**V.** To analyse the Normality of Distribution of the developmental sample.

**VI.** To analyse items of drafts of ELPT using “Cronbach’s Alpha if Item Deleted” to establish reliability.

**VII.** To analyse the Cronbach’s Alpha of the final draft of ELPT to establish its reliability.

These objectives that aided in the standardization of the ELPT were also constitutive of the process of its very construction, the details of which now follow.

### **A. Skill-wise six-point Grading scale to assess English Language skills at Elementary level.**

As a part of developing the English Language Proficiency Test, a system of grading for grading/scoring the proficiency qualitatively and quantitatively was required to be developed. The language skills were defined in operational terms, identifying the most important parameters for their assessment, and a grading scale was developed for this assessment. Initially, a 10-point grading scale was developed and experts’ feedback was obtained, following which the scale was modified accordingly for scoring the initial draft of ELPT, and at the final stage, a 6-point grading scale was retained.

#### **A.1. Listening Skill Grading Scale**

**Operational Definition of Listening:** ‘Listening Skill’, in this study, means “hearing with comprehension”. In other words, ‘Listening’ incorporates an understanding of what is being heard.

Listening is an active skill, for it is a voluntary act unlike ‘Hearing’ which is a passive and involuntary act: ‘listening’ requires attention and a certain level of ability to focus on what is being heard as opposed to ‘hearing’, which requires neither of these.

Listening is, thus, successful when the listener is able to understand - comprehend - the meaning of the words being listened to and can, if required, meaningfully convey what has been said or respond verbally or/and non-verbally to it. Other than attention and ability to focus, it would also require a workable knowledge of the vocabulary and grammar of the language being listened to, for without that comprehension will most likely be incomplete.

#### **Assessment Parameters of Listening Skill**

1. Whether the student was able to correctly understand what was being listened to in a minimum number of attempts without missing out on any important detail.
2. Whether the student was able to give the appropriate response to the questions related to the material listened to.
3. Whether the student was able to understand the vocabulary of the matter being listened to.

**Grading points:** Taking into account all the Assessment Parameters, the Level of Listening Proficiency can be measured and graded on the following **Six-point Grading scale** in which a Higher Numeric Grade Point corresponds to the Higher Proficiency in **Listening Skill**.

**Grade 1:** The student could not understand audio even after it was played out **thrice** and gave **no/incorrect** verbal/non-verbal response, **even after receiving a clue or gave a completely unrelated answer**.

**Grade 2:** The student could not understand the audio and **didn't ask for a replay or for a clue** but gave an **almost incorrect** verbal/non-verbal response missing out on most of the crucial information except one or two small pieces of information.

**Grade 3:** The student understood the audio after it was played out **thrice** and gave **almost correct** verbal/non-verbal response, **after receiving a clue but missed out on more than one crucial information/word-meaning**.

**Grade 4:** The student understood the audio after it was played out **twice** and, **after receiving a clue** gave the **correct** verbal/non-verbal response but **missed out on just one crucial information/word-meaning**.

**Grade 5:** The student understood the audio after it was played out **once** and gave a **correct** response, **without asking for/ receiving a clue but missed out on a crucial information/word-meaning**.

**Grade 6:** The student understood the audio after it was played/spoken out **once** and gave the **correct** response confidently **without receiving a clue**.

## **A.2 Speaking Skill Grading Scale**

**Operational Definition:** 'Speaking Skill', in this Study, means the ability to convey intended information, ideas, thoughts or feelings 'intelligibly' in required word-length. 'Intelligibly' means that the matter was spoken with clear articulation and a certain level of smoothness/fluency using the appropriate words, phrases and with the absence of too many grammatical lapses while following a somewhat logical sequence.

This definition attempts to incorporate a very important insight about the purpose and criterion for the assessment of the Speaking Skill as given by the Position Paper on English by NCERT thus: "Discussions on what model of English is appropriate for India centre mostly around the pronunciation (i.e. spoken English), and secondarily on lexis or vocabulary. The criterion for an acceptable pronunciation has to be intelligibility."

In this definition, the exactness of pronunciation, tone, tenor and accent may not perhaps be of great importance, the main criterion being ease of communication with the purpose of facilitating understanding and comprehension of what is being said.

Here, 'Speaking' is different from 'Public Speaking' i.e. delivering speeches or monologues or 'Mass-Communication' through Radio, Television etc. which are based on one-way communication and are therefore "non-interactive" in nature. In this Study, 'Speaking' has been taken only in the "interactive" or "partially interactive" formats, such as conversations or describing one's locality.

## Assessment Parameters of Speaking Skill

Taking into account all the Assessment parameters, the Level of Speaking Proficiency can be measured and graded on the following **Six-point Grading scale** in which Higher Numeric Grade Point corresponds to the Higher Proficiency in **Speaking Skill**.

- Whether the student was able to convey the intended information, ideas, thoughts or feelings with clear articulation.
- Whether the student was able to speak smoothly or fluently with proper intonation.
- Whether the student was able to choose appropriate words and phrases for what (s)he wanted to convey in spoken form.
- Whether the student was able to frame 'intelligible' sentences without too many grammatical lapses that might have diluted the understanding of the intended message for the listener.
- Whether the student was able to follow a somewhat logical sequence while conveying intended information, ideas, thoughts or feelings.

## Grading Points

- **Grade 1: No Response/Wrong Response/Completely unrelated response to the question asked/ Unintelligible response**
- **Grade 2:** The student's verbal response of a **word/phrase** was '**intelligible**' which was also in consonance with the question asked/demand of the situation but **didn't require much proficiency in English** as the response was that of some factual information, such as name, place, date etc. Or, **in a long series of spoken sentences**, the grammatical mistakes and/or lack of clarity in articulation/lack of fluency (if any) **hindered the intelligibility** of the response to a great extent.
- **Grade 3:** The student spoke only a **complex sentence** or **short series of sentences in consonance with the question asked** or demand of the situation. Or, in a long series, the grammatical mistakes and/or lack of clarity in articulation/lack of fluency **hindered the intelligibility** of the sentences. The sentences did not follow a logical sequence throughout. Or the response was **too short i.e. about 1/3<sup>rd</sup>** of the number of sentences asked to speak.
- **Grade 4:** The student spoke a **series of sentences in consonance with the question asked** or demand of the situation. The grammatical mistakes and/or lack of clarity in articulation/lack of fluency **didn't much hinder the intelligibility** of the sentences. The sentences did not follow a logical sequence throughout. Or the response was **short i.e. about half** of the number of sentences asked to speak.
- **Grade 5:** The student spoke a series of sentences in consonance with the question asked or demand of the situation. The grammatical mistakes and/or lack of clarity in articulation/lack of fluency **didn't hinder the intelligibility** of the sentences. The sentences didn't follow a **logical sequence** throughout. The response was **short i.e. about 2/3<sup>rd</sup>** of the number of sentences asked to speak. There was a lack of appropriateness in the word choice in terms of choosing wrong words of English or occasional use of words from Hindi or Mother tongue.

➤ **Grade 6:** The student spoke a series of sentences in consonance with the question asked or demand of the situation. There were **no grammatical mistakes** or they were negligible. The articulation was clear and **fluent**. The sentences followed a **logical sequence** throughout. The response was **full length** i.e. had the number of sentences the student was asked to speak. The **word-choice was appropriate**.

### **A.3 Reading Skill Grading Scale**

**Operational Definition:** ‘Reading’ is not just decoding i.e. deciphering the words and sentences being read. It would include doing so with a certain level of fluency accompanied by an understanding of the meaning of the information being conveyed as also the ability to draw inferences from what is being read which, in general, is termed as Reading Comprehension. Reading skill, in this study incorporates both the Reading Aloud ability and Reading Comprehension. A good Reading Aloud exercise will be that in which text is read aloud with minimum omissions, substitutions, insertions and mispronunciations.

The Focus Group Position Paper on English by NCERT identifies the following sub-skills of the Reading Skill:

- a) “Reading Aloud/ Decoding: As children become more proficient in decoding, they read faster (words per minute) with fewer mistakes.
- b) Scanning a text (such as a list, a telephone directory, an advertisement) for information
- c) Reading for given information (factual comprehension)
- d) Reading for inference
- e) Extended reading”

The present study does not incorporate the ‘Extended Reading’ sub-skill.

### **Assessment Parameters for Reading Skill**

➤ The student, while Reading Aloud, was able to decode words and sentences with minimum omissions, substitutions, insertions and mispronunciations. (S)he was able to maintain an optimum pace of reading in order to facilitate understanding for herself while reading.

- The student was able to grasp the factual information from the text.
- The student was able to understand the theme of the entire text.
- The student was able to understand and compare different ideas and thoughts.
- The student was able to draw inferences from what was being read in terms of thrust of the arguments being put forward.
- The student was able to provide a suitable title to the paragraph after reading.

### **Grading Scale of Reading Skill**

Taking into account all the Assessment parameters, the Level of Reading Proficiency can be measured and graded on the following **Six-point Grading scale** in which Higher Numeric Grade Point corresponds to the Higher Proficiency in **Reading Skill**. Reading



Skill has been divided into **two parts i.e. Reading Aloud Ability (RAA) and Reading Comprehension** for which separate grading scales have been developed.

The Grading Scale for **Reading Aloud Ability** is given below

**Note:** The word count for Reading Aloud Ability is reached after subtracting the Omissions of words from the written text (O), Substitutions of the actual words in the text (S), Insertions that were made in the original text (I) and Mispronunciations {(M) (not counted if student corrects it without having been given a clue)}.

**RAA= Words Read per minute (W)- (O+S+I+M)**

**In terms of pace and fluency of reading, the performance is considered to be of the Optimum, Far Above Satisfactory, Above Satisfactory, Satisfactory, Below Satisfactory and Far Below Satisfactory levels and these have been categorized into Grades thus :**

**The Grades are defined as below.**

- **Grade 1:** Reading Aloud ability is in the range of 0-20 words per minute. Pace and fluency while reading can be termed as 'Far Below Satisfactory Level'.
- **Grade 2:** Reading Aloud ability is in the range of 21-40 words per minute. Pace and fluency while reading can be termed as 'Below Satisfactory Level'.
- **Grade 3:** Reading Aloud ability is in the range of 41-60 words per minute. Pace and fluency while reading can be termed as 'Satisfactory Level'.
- **Grade 4:** Reading Aloud ability is in the range of 61-80 words per minute. Pace and fluency while reading can be termed as 'Above Satisfactory Level'.
- **Grade 5:** Reading Aloud ability is in the range of 81-100 words per minute. Pace and fluency while reading may be termed as 'Far Above Satisfactory Level'.
- **Grade 6:** Reading Aloud ability is more than 100 words per minute. 'Optimum' pace, fluency and intonation were maintained while reading.
- 

**Reading Comprehension Grading Scale**

**Grading was done on the basis of the student's level of grasp of factual information, theme, ability to draw inferences and give a title to a passage/text etc.**

**Grade 1:**No response/ Wrong Response/ Completely Unrelated Response.

**Grade 2:** The student **attempted but could neither grasp the factual information, ideas, thoughts etc. except a very prominent name/date etc.** from the text nor he could compare the given ideas or draw inferences and get hold of the theme. (S)he was unable to provide a suitable title for the text.

**Grade 3:** The student could grasp **some major pieces of factual information** from the text but (S)he could neither understand the ideas, thoughts etc nor compare the given ideas or draw inferences and get hold of the theme. (S)he was unable to provide a suitable title for the text.

**Grade 4:** The student could grasp all the pieces of factual information from the text and he could also understand the ideas and thoughts as well as compare and contrast the

given ideas while drawing inferences but **could not get hold of the theme**. (S)he was also **unable to provide a suitable title for the text**.

**Grade 5:** The student could grasp all the pieces of factual information from the text and (s)he could also understand the ideas and thoughts as well as compare and contrast the given ideas while drawing inferences and getting hold of the theme but was **unable to provide the suitable title of the passage**.

**Grade 6:** The student could grasp all the pieces of factual information from the text and (s)he could also understand the ideas and thoughts as well as compare and contrast the given ideas while drawing inferences and getting hold of the theme as well as the **title of the passage**.

#### **A.4 Writing Skill Grading Scale**

**Operational Definition:** The “Writing Skill”, in this study, is the ability to express one’s feelings, thoughts or a piece of information in the form of written text which is intelligible, somewhat organized and broadly conforms to the niceties of grammatical structures. This means that successful writing would require a working knowledge of spellings, age/level-appropriate vocabulary, and grammar along with the ability to plan and produce a coherent piece of writing. As the writing skill develops, students gradually evolve from single-word responses to simple sentences to elaborate stories and essays, using more appropriate words and improved grammatical structures, better coherence, lesser spelling mistakes, correct word order and minimal lapses in punctuation. Writing skill, in this study, doesn’t take into account the beauty or neatness of handwriting as long as it is ‘intelligible’.

#### **Assessment Parameters of Writing Skill**

1. Whether the student was able to express her/himself in the form of a written text in a manner that, to a reasonable extent, conforms to the niceties of grammatical structures.
2. Whether the student was able to express her/himself in the written form that reflects a working knowledge of spellings and age/level-appropriate vocabulary.
3. Whether the student was able to produce a fairly intelligible written text with correct word order and minimal lapses in punctuation.
4. Whether the student, while expressing her/himself in writing, was able to produce a coherent piece of composition, reflecting a somewhat logical progression of thoughts and ideas, with an appropriate beginning, middle and end.

#### **Grading Points**

Taking into account all the Assessment Parameters, the Level of Writing Proficiency can be measured and graded on the following **Six-point Grading Scale** in which Higher Numeric Grade Point corresponds to the Higher Proficiency in **Writing Skill**:

**Grade 1: No Response/Wrong Response/Completely unrelated response to the question asked/ Unintelligible response**

**Grade 2:** The student's written response of a **Word/phrase/simple sentence** was '**intelligible**' which was also in consonance with the question asked. Or in a series of sentences, there were **too many spelling and/or grammatical mistakes** or answer wasn't in consonance with the question asked. The student depicted below satisfactory knowledge of punctuation marks and word order in a sentence (3 or more mistakes in punctuation and word order collectively).

**Grade 3:** The student's written response of only a **complex sentence** or **short series of sentences** was in consonance with the question asked. Or, in a long series, the grammatical mistakes and/or spelling mistakes/wrong usage of punctuation marks **hindered the intelligibility** of the sentences to some extent. The student depicted **satisfactory knowledge of punctuation marks and word order in a sentence** (In a sentence, not more than two mistake in word order and punctuation collectively).

**Grade 4:** The student wrote a **series of sentences** in consonance with the question asked. The spelling/grammatical mistakes **didn't much hinder the intelligibility** of the sentences. The sentences **did not follow a logical sequence** throughout. The response was **short i.e. about half** of the number of sentences asked to write. In arranging jumbled sentences, student depicted good knowledge of punctuation marks and word order in a sentence. (No or negligible mistakes in punctuation and word order collectively).

**Grade 5:** The student wrote a series of sentences in consonance with the question asked. The spelling/ grammatical mistakes **didn't hinder the intelligibility** of the sentences. The sentences **didn't follow a logical sequence** throughout. The response was **short i.e. less than 2/3<sup>rd</sup>** of the number of sentences asked to write. There was a **lack of appropriateness in the word choice** in terms of choosing wrong words of English or occasional use of words from Hindi or Mother tongue. The student depicted above satisfactory knowledge of punctuation marks and word order in a sentence. (No or negligible mistakes in punctuation and word order collectively).

**Grade 6:** The student wrote a series of sentences in consonance with the question asked. There were **no spelling/grammatical mistakes** or they were negligible. The sentences followed a logical sequence throughout. The response was **full length** i.e. had the number of sentences the student was asked to write. The **word choice was appropriate**. The student depicted above satisfactory knowledge of punctuation marks and word order in a sentence. (No or negligible mistakes in punctuation and word order collectively)

## **B. Resources Consulted (Inclusive of Interviewing Experts)**

Good planning for a test includes a number of sub-tests, test items and proper components of the content to be dealt with. Just to be more familiar with all of these, the researcher went through related books, documents from NCERT such as "Learning Indicators and Learning Outcomes at the Elementary Stage" (2014), National Curriculum Framework, Position Paper on Teaching of English, and tests like TOEFL and IELTS, published theses and dissertations related with this field. The researcher also studied some books of language, language-testing and assessment of proficiency in the

English language. In addition to this, the researcher interviewed 14 well-known educationists, experts of test-construction, experts of pedagogy of English and senior teachers of elementary schools.

### **C. Preparing a pool of Questions and Blueprint**

Taking insights from the experts' interviews, some of the learning indicators prescribed by NCERT in its document "Learning Indicators and Learning Outcomes at the Elementary Stage" (2014), some learning indicators from all the skills were identified as being more important. They were selected for developing questions around them, as it was not possible to develop content for all of the learning indicators in a single package. The researcher prepared a pool of 72 questions covering all the language skills i.e. listening, speaking, reading and writing. All the inputs taken from related literature and interviews of experts helped the researcher in determining the test-content.

### **D. Gaining the Opinion of Experts and Preliminary Modification**

This being a second-language test, the content pertaining to English as a second language must be determined keeping in mind that language is a skill-subject and assessment of language proficiency poses multi-layered challenges. Bearing in mind the importance of experience in the field, the researcher tried out the test based on the prepared pool of 72 questions on 20 experts of the English language and their review was obtained in order to remove ambiguous items and to detect defects in the format and/or language, if any, or vagueness of the items. Therefore, from the pool of 72 questions prepared by the researcher, 12 questions were removed, considering the experts' opinion and 60 questions were retained for the first draft of ELPT.

### **E. Item Analysis**

Item analysis is a statistical procedure undertaken for selection and rejection of the items of a test on the basis of their difficulty-value and discriminative value. It is undertaken to obtain the best items of the test.

"When norm-referenced tests are developed for instructional purposes, to assess the effects of educational programs, or for educational research purposes, it can be very important to conduct item and test analyses. These analyses evaluate the quality of the items and of the test as a whole. Such analyses can also be employed to revise and improve both items and the test as a whole" (Hetzl, 1997).

It is clear from this reference that one of the most crucial stages of test-construction, as a part of the assessment of the test, is Item Analysis. It is meant for "testing the test" in order to identify the unsuitable test-items and also, to zero in on the tedious content which causes the students trouble. Item analysis is done by equating the group of students in the Upper Criterion Group (U) and Lower Criterion Group (L). These groups are selected on the basis of the criteria suggested by (Kelley, 1939), according to which "In a normal distribution, the ideal point when these two conditions offset is 27 percent". Therefore, scores obtained by the students on the Initial draft of ELPT were arranged in descending order in MS Excel and 27% students with the highest scores formed the

Upper criterion group (U) and 27% students with the lowest scores formed the Lower criterion group (L). As the scale was administered to 120 students, 32 students formed the upper and lower criterion group keeping all the cut-off point scorers within the group. As the format of the ELPT is in the form of grading scale in which proficiency in language-skills is assigned grades in increasing order of numerals as per the proficiency of a student and there were multiple correct responses possible, Difficulty Value cannot be calculated because of the possibility of more than one correct responses out of all the responses given. Therefore, finding out the Discriminatory Power of each item, was considered pertinent for assessing the suitability of the items. This was done by administering t-test to check the significance of the difference between upper and lower criterion group for each item as presented in Table 4.3

**Table 4.4 t-value for each Item in Initial Draft of Proficiency Test**

Item No.	Pair No.	Mean	S. D.	Std. Error Mean	t-value	df	Sig. (2-tail)
1	1(L)-1(U)	1.56250	1.01401	.17925	8.717	31	.000
2	2(L) -2(U)	.18750	1.22967	.21738	.863	31	.395
3	3(L) -3(U)	2.87500	1.23784	.21882	13.139	31	.000
4	4(L) -4(U)	.50000	1.19137	.21061	2.374	31	.024
5	5(L) -5(U)	2.21875	.75067	.13270	16.720	31	.000
6	6(L) -6(U)	.18750	1.99091	.35195	.533	31	.598
7	7(L) - 7(U)	3.00000	2.48868	.43994	6.819	31	.000
8	8(L) - 8(U)	-.21875	1.92997	.34117	-.641	31	.526
9	9(L) - 9(U)	3.09375	2.44104	.43152	7.169	31	.000
10	10(L)-10(U)	-.90625	1.48887	.26320	-3.443	31	.002
11	11(L) - 11(U)	.18750	2.53285	.44775	.419	31	.678
12	12(L) - 12(U)	2.96875	2.30685	.40780	7.280	31	.000
13	13(L) - 13(U)	.43750	1.81281	.32046	1.365	31	.182
14	14(L) - 14(U)	1.21875	1.49697	.26463	4.605	31	.000
15	15(L) - 15(U)	1.90625	2.16064	.38195	4.991	31	.000
16	16(L) - 16(U)	2.53125	2.06326	.36474	6.940	31	.000

<b>17</b>	17(L) - 17(U)	1.37500	.55358	.09786	14.051	31	<b>.000</b>
<b>18</b>	18(L) - 18(U)	-.25000	.98374	.17390	-1.438	31	<b>.161</b>
<b>19</b>	19(L) - 19(U)	.68750	.78030	.13794	4.984	31	<b>.000</b>
<b>20</b>	20(L) - 20(U)	3.18750	2.33401	.41260	7.725	31	<b>.000</b>
<b>21</b>	21(L) - 21(U)	.09375	1.14608	.20260	.463	31	<b>.647</b>
<b>22</b>	22(L) - 22(U)	.28125	1.95488	.34558	.814	31	<b>.422</b>
<b>23</b>	23(L) - 23(U)	2.96875	2.55878	.45233	6.563	31	<b>.000</b>
<b>24</b>	24(L) - 24(U)	-.34375	.78738	.13919	-2.470	31	<b>.019</b>
<b>25</b>	25(L) - 25(U)	.21875	1.64090	.29007	.754	31	<b>.456</b>
<b>26</b>	26(L) - 26(U)	.62500	.94186	.16650	3.754	31	<b>.001</b>
<b>27</b>	27(L) - 27(U)	3.18750	2.37511	.41986	7.592	31	<b>.000</b>
<b>28</b>	28(L) - 28(U)	.53125	2.28578	.40407	1.315	31	<b>.198</b>
<b>29</b>	29(L) - 29(U)	2.12500	1.09985	.19443	10.929	31	<b>.000</b>
<b>30</b>	30(L) - 30(U)	-.03125	1.76862	.31265	-.100	31	<b>.921</b>
<b>31</b>	31(L) - 31(U)	1.53125	2.04757	.36196	4.230	31	<b>.000</b>
<b>32</b>	32(L) - 32(U)	1.09375	1.85541	.32799	3.335	31	<b>.002</b>
<b>33</b>	33(L) - 33(U)	1.75000	1.75977	.31109	5.625	31	<b>.000</b>
<b>34</b>	34(L) - 34(U)	.65625	1.20775	.21350	3.074	31	<b>.004</b>
<b>35</b>	35(L) - 35(U)	.43750	1.36636	.24154	1.811	31	<b>.080</b>
<b>36</b>	36(L) - 36(U)	.56250	1.10534	.19540	2.879	31	<b>.007</b>
<b>37</b>	37(L) -	2.62500	2.53683	.44845	5.853	31	<b>.000</b>

	37(U)						
<b>38</b>	38(L) - 38(U)	.37500	1.69915	.30037	1.248	31	<b>.221</b>
<b>39</b>	39(L) - 39(U)	.53125	.67127	.11867	4.477	31	<b>.000</b>
<b>40</b>	40(L) - 40(U)	2.00000	1.72271	.30454	6.567	31	<b>.000</b>
<b>41</b>	41(L) - 41(U)	.59375	1.01153	.17881	3.320	31	<b>.002</b>
<b>42</b>	42(L) - 42(U)	.53125	1.19094	.21053	2.523	31	<b>.017</b>
<b>43</b>	43(L) - 43(U)	3.06250	2.42218	.42818	7.152	31	<b>.000</b>
<b>44</b>	44(L) - 44(U)	-.21875	1.03906	.18368	-1.191	31	<b>.243</b>
<b>45</b>	45(L) - 45(U)	2.43750	1.91661	.33881	7.194	31	<b>.000</b>
<b>46</b>	46(L) - 46(U)	1.87500	1.80947	.31987	5.862	31	<b>.000</b>
<b>47</b>	47(L) - 47(U)	1.28125	1.17045	.20691	6.192	31	<b>.000</b>
<b>48</b>	48(L) - 48(U)	2.03125	1.84013	.32529	6.244	31	<b>.000</b>
<b>49</b>	49(L) - 49(U)	1.93750	.75935	.13424	14.434	31	<b>.000</b>
<b>50</b>	50 (L)- 50(U)	.43750	.75935	.13424	3.259	31	<b>.003</b>
<b>51</b>	51(L) - 51(U)	0.00000	1.58623	.28041	0.000	31	<b>1.000</b>
<b>52</b>	52(L) - 52(U)	1.46875	1.26960	.22444	6.544	31	<b>.000</b>
<b>53</b>	53(L) - 53(U)	1.56250	1.56447	.27656	5.650	31	<b>.000</b>
<b>54</b>	54(L) - 54(U)	1.43750	1.68365	.29763	4.830	31	<b>.000</b>
<b>55</b>	55(L) - 55(U)	2.31250	1.89119	.33432	6.917	31	<b>.000</b>
<b>56</b>	56(L) - 56(U)	2.93750	2.36831	.41866	7.016	31	<b>.000</b>
<b>57</b>	57(L) - 57(U)	1.56250	2.52647	.44662	3.498	31	<b>.001</b>

<b>58</b>	58(L) - 58(U)	.03125	.47413	.08381	.373	31	<b>.712</b>
<b>59</b>	59(L) - 59(U)	3.15625	2.28755	.40438	7.805	31	<b>.000</b>
<b>60</b>	60(L) - 60(U)	<b>.21875</b>	<b>.75067</b>	<b>.13270</b>	<b>1.648</b>	<b>31</b>	<b>.109</b>

As an item should be accepted if it has significant value or p-value less than 0.05 (i.e.  $p < .05$ ), from table 4.4, it is clear that sig. value of all the items except item no. 2,6,8,11,13,18,21,22,25,28,30,35,38,44,51,58 and 60 is less than 0.05. So, these seventeen items (2,6,8,11,13,18,21,22,25,28,30,35,38,44,51,58 and 60) were rejected and excluded from the first draft. The remaining items were rearranged in serial order from 1 to 43. The second draft now contained 43 items (questions). Cronbach's Alpha for the first draft of ELPT is given in the table below.

**Table- 4.5 Cronbach's Alpha for the first draft of ELPT**

<b>Cronbach's Alpha</b>	<b>N of Items</b>
<b>.898</b>	<b>60</b>

#### **F. Seeking the Opinions of Experts for Final Draft**

The retained questions in the ELPT were reviewed for content legitimacy by various language experts, test-construction experts and other specialists from various teaching subjects because "Acquiring content approval is additionally an essential part in the scale improvement process" (DeVellis, 2003). They reviewed the content legitimacy of items (questions) of ELPT in terms of appropriateness of language, suitability as a measure of concerned skill on a 3-point Likert scale: (0: completely unsuitable, 1: partially suitable 2: totally suitable). On the basis of review obtained from them, some amendments were made in the second draft of ELPT.

**Table: 4.6 Item total correlation for the second draft of ELPT**

<b>Item</b>	<b>Scale Mean if Item Deleted</b>	<b>Scale Variance if Item Deleted</b>	<b>Corrected Item-Total Correlation</b>	<b>Cronbach's Alpha if Item Deleted</b>
<b>14</b>	141.6750	1012.356	.876	.935
<b>06</b>	141.6083	1012.761	.863	.935
<b>30</b>	141.6083	1013.350	.858	.935
<b>43</b>	141.7333	1014.382	.857	.935
<b>15</b>	141.6750	1014.826	.856	.935



<b>11</b>	141.5000	1033.866	.851	.935
<b>05</b>	141.6917	1015.560	.849	.935
<b>08</b>	141.6833	1017.798	.848	.935
<b>27</b>	141.3167	1049.832	.847	.936
<b>41</b>	141.7333	1017.172	.840	.935
<b>34</b>	141.3667	1048.789	.827	.936
<b>32</b>	141.3250	1053.230	.814	.936
<b>04</b>	143.8917	1069.845	.805	.937
<b>22</b>	141.3083	1060.097	.800	.936
<b>25</b>	141.5667	1028.298	.764	.936
<b>35</b>	144.2833	1080.675	.754	.937
<b>18</b>	142.2167	1021.785	.735	.936
<b>12</b>	144.7750	1096.714	.626	.938
<b>19</b>	143.5667	1077.844	.604	.938
<b>02</b>	144.1833	1067.411	.600	.937
<b>23</b>	144.2417	1099.597	.520	.938
<b>20</b>	141.8333	1081.485	.499	.938
<b>37</b>	143.2333	1093.659	.497	.938
<b>01</b>	144.9583	1104.645	.389	.939
<b>26</b>	143.8667	1117.310	.337	.939
<b>33</b>	143.7167	1108.423	.307	.939
<b>24</b>	143.2667	1109.575	.286	.940
<b>39</b>	143.5917	1102.176	.286	.940
<b>38</b>	143.6583	1101.302	.285	.940
<b>40</b>	143.0333	1095.747	.264	.940
<b>13</b>	144.0167	1117.764	.250	.940
<b>31</b>	143.5500	1094.855	.247	.941

28	143.9583	1115.671	.245	.940
29	144.1667	1116.325	.198	.940
10	142.7833	1111.028	.164	.941
17	143.9167	1121.254	.142	.940
36	145.1667	1125.115	.105	.940
09	143.7417	1119.924	.102	.941
42	144.3583	1113.425	.095	.942
03	144.9583	1124.393	.069	.941
21	142.4083	1123.689	.039	.942
16	145.2667	1144.046	-.378	.941
07	144.9500	1171.510	-.581	.944

It is recommended by Hanover (2010), that if an item has item total correlation near to zero, then it may be considered for removal as it is not measuring the same thing as the rest of the items of the test. Therefore, to improve the reliability of ELPT, out of 43 items, 7 items (3, 7, 9, 16, 21, 36, 42) were deleted (reducing the number of items now to 36) as either their correlation was negative or near to 0.

**Table 4.7 item-total correlation for the second draft of ELPT after first elimination**

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
11	126.9667	1032.537	.928	.951
05	126.9000	1032.292	.920	.951
25	126.9000	1032.881	.915	.952
12	126.9667	1034.352	.913	.952
36	127.0250	1034.209	.912	.952
04	126.9833	1035.159	.905	.952
06	126.9750	1037.722	.903	.952
22	126.6083	1072.526	.896	.952

08	126.7917	1056.116	.895	.952
35	127.0250	1037.235	.894	.952
29	126.6583	1071.386	.875	.952
17	126.6000	1082.225	.862	.953
27	126.6167	1076.474	.857	.953
20	126.8583	1047.786	.824	.952
03	129.1833	1097.210	.805	.953
14	127.5083	1042.235	.782	.953
30	129.5750	1108.583	.747	.954
09	130.0667	1125.441	.606	.955
15	128.8583	1105.719	.599	.954
02	129.4750	1096.941	.578	.954
18	129.5333	1124.856	.564	.955
16	127.1250	1108.060	.509	.955
31	128.5250	1124.571	.451	.955
01	130.2500	1135.618	.339	.956
21	129.1583	1145.496	.330	.956
19	128.5583	1135.761	.311	.956
28	129.0083	1139.958	.262	.956
32	128.9500	1133.073	.253	.956
23	129.2500	1144.508	.229	.956
33	128.8833	1136.087	.218	.956
10	129.3083	1148.232	.196	.956
24	129.4583	1145.326	.182	.956
34	128.3250	1136.591	.153	.958
26	128.8417	1136.538	.137	.958
13	129.2083	1150.671	.117	.956

07	128.0750	1150.356	.052	.958
----	----------	----------	------	------

Table- 4.7 depicts that the 36 items of ELPT were again rearranged and tested for item-total correlation. 8 items (7, 10, 13, 23, 24, 26, 33, 34) having correlation below .25 were removed to improve the reliability of the test, thereby further bringing the total down to 28.

**Table 4.8 Cronbach's Alpha after second elimination**

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
01	101.5583	1027.576	.288	.976
02	100.7833	994.440	.507	.975
03	100.4917	989.983	.777	.974
04	98.2917	922.578	.962	.972
05	98.2083	919.713	.978	.972
06	98.2833	925.364	.958	.972
07	98.1000	945.217	.933	.973
08	101.3750	1019.631	.510	.975
09	98.2750	921.041	.977	.972
10	98.2750	921.579	.972	.972
11	98.8167	930.302	.826	.974
12	100.1667	1001.703	.528	.975
13	98.4333	998.382	.508	.975
14	97.9083	968.756	.927	.973
15	100.8417	1011.899	.609	.975
16	99.8667	1022.755	.341	.976
17	98.1667	934.190	.884	.973
18	100.4667	1035.024	.299	.976
19	97.9167	960.296	.948	.973
20	98.2083	920.150	.974	.972

21	97.9250	964.204	.907	.973
22	97.9667	959.058	.927	.973
23	100.8833	1003.734	.667	.975
24	99.8333	1022.157	.330	.976
25	98.3333	924.073	.955	.972
26	98.3333	922.106	.965	.972
27	98.1000	945.217	.933	.973
28	98.1667	934.190	.884	.973

**Overall Chronbach alpha** of the test was found to be **.972**. As shown in the table-4.8, Chronbach's alpha 'if item deleted' for every item is more than the **overall Chronbach alpha i.e. 0.972**. Therefore, all the items were found to be reliable and hence accepted. The acceptable value of reliability coefficient as suggested by Nunnally (1978) is 0.7. Therefore, it can be said that ELPT has a sufficiently high level of reliability.

#### **Another Round of Seeking the Opinions of Experts for Final Draft**

In this way, final draft of ELPT was obtained which was again sent to be reviewed for content-legitimacy by various language experts, test-construction experts and other specialists from various teaching subjects, given the fact that "Acquiring content approval is additionally an essential part in the scale improvement process" (DeVellis, 2003). They again reviewed the items (questions) and ELPT as a whole in terms of content, language and quality of being a sufficient measure of proficiency in the English language. On the basis of the review obtained from them, some amendments were made in the final draft of ELPT.

#### **G. Final Run**

To take a final decision on the items in a scale, it needs to be administered to a substantially large sample of subjects. So, considering the form of the final draft of ELPT, which requires to be administered to an individual student, it was administered to 120 randomly selected students of 8<sup>th</sup> class of Govt. schools of Mahendergarh district of Haryana.

Before administering the English Language Proficiency Test, all the necessary preparations were made accordingly. These preparations included ensuring the functioning of equipments and careful observation of the forms of the ELPT. Collected data was entered into Microsoft Excel spreadsheet and then its Mean (185.12), Median (187.52), Skewness (-.537), and Kurtosis (-.031) values were calculated.

#### **Table 4.9 Descriptive Statistics of ELPT Score**

Descriptives				
			Statistic	Std. Error
Obtained score	Mean		185.1167	2.86484
	95% Confidence Interval for Mean	Lower Bound	179.4440	
		Upper Bound	190.7893	
	5% Trimmed Mean		186.5185	
	Median		187.0000	
	Variance		984.877	
	Std. Deviation		31.38275	
	Minimum		98.00	
	Maximum		237.00	
	Range		139.00	
	Interquartile Range		45.75	
	Skewness		-.537	.221
	Kurtosis		-.031	.438

Table 4.9 above depicts that the Skewness and Kurtosis falls within the acceptable range (i.e. Skewness < |2.0| and Kurtosis < |9.0|; (Posten, 1984) which means that the sample is normally distributed.

The final form of ELPT consisted of 28 items and the complete ELPT was tested for its reliability by once again calculating Cronbach's Alpha coefficients and Item-Total Correlations. Overall **Cronbach's Alpha was found to be .972** for all the items included in the finalized version of ELPT. The reliability coefficients indicated that overall scale had acceptable internal consistency.

Reliability being one of the most important characteristics of any test, it was measured through another method which is called "Test-Retest Method". "Reliability means consistency of scores obtained by same individual when re-examined with test on different sets of equivalent items or under other variable examining condition" Anastasi (1968). Keeping in view this recommendation, the test was administered to the same group of students after fifteen days and two sets of scores were correlated. The **reliability co-efficient** of ELPT through this method was found to be **.86** which confirms the **high reliability** of the test.

The final form of ELPT has the following features:

1. Total number of questions is 28 (twenty eight). 6 questions have been assigned to Listening skill; Speaking skill has 7 questions; Reading skill, 8 questions; and Writing skill, 8 questions.
2. Equal distribution of weightage to all the four skills i.e. Listening, Speaking,

Reading and Writing i.e. 36 marks/grades to every skill.

3. Questions on Listening and Reading skills are multiple choice questions in which every answer has been assigned different grades depending upon the level of proficiency it depicted in a skill.

4. Questions on Speaking and Writing skill seek descriptive answers; therefore, they cannot be framed in multiple choice forms. Answers to these questions are assigned grades following their respective grading scales. In order to reduce the element of subjectivity in assigning grades, the answers were checked by three teachers of the English language and the nearest whole number to the average numeric grade was assigned to the answer.

## H. Calculating Validity and Reliability

### Validity of the test

(Ghiselli, 1964) defined validity of a test as; “the extent to which a test or a set of operations measures what it is supposed to measure”.

In order to ascertain that the ELPT measures what it is supposed to measure, its content validity was reviewed through expert opinion. Content validity is an important criterion for ensuring the usefulness of the test, especially of a proficiency test. It was observed by the experts that the items of the ELPT were distributed over majority of the instructional objectives and assessed all the 4 language skills in a holistic manner. Thus, review from subject matter experts confirmed the validity of the test.

---

### References

Anastasi, A. (1968). Psychological testing (3<sup>rd</sup> Ed.). Macmillan: The University of Michigan.

De Vellis, R. F. (2003). Scale Development: Theory and Applications (2nd ed., Vol. 26). Thousand Oaks, CA: Sage Publications.

Ghiselli, H.H. (1964). Theory of Psychological Measurement. New York: McGraw Hill.

Hanover (2010). Reliability Analysis. Retrieved from <https://psych.hanover.edu/classes/ResearchMethods/Assignments/reliability-1.html>

Hetzl, S. M. (1997). Basic concepts in item and test analysis. Texas A&M University. Retrieved from <http://ericae.net/ft/tamu/Espy.htm>

NCERT (2006). National Focus Group Position Paper on Teaching of English. Retrieved from

[http://www.ncert.nic.in/new\\_ncert/ncert/rightside/links/pdf/focus\\_group/english.pdf](http://www.ncert.nic.in/new_ncert/ncert/rightside/links/pdf/focus_group/english.pdf)

NCERT (2014). Learning Indicators and Learning Outcomes at the Elementary Stage. Retrieved from <http://www.dsek.nic.in/Misc/learningoutcome.pdf>

Nunnally, J.C. (1972). Educational measurement and evaluation (2nd ed.). New York: McGraw-Hill.

Posten, H.O. (1984). Robustness of the two-sample t-test. In D. Rasch and M.L. Tiku (Eds.). Robustness of Statistical Methods and non-parametric statistics. Dordrecht, German Democratic Republic: D. Reidel.