

Background

Based on growing international needs for the English language, several academic and non-academic institutions approached the National Center for Assessment in Higher Education calling for the development of an English test that could measure the proficiency of their applicants. Consequently the Center formed a committee consisting of specialists in the fields of English and psychological measurement. The Committee looked first into available international, regional and local test experiences and practices. It also sought the consultation of institutions that specialize in standardized language testing and even made field visits to some of these institutions. After the survey the Committee reached the conclusion that there was considerable need for a test that could fulfill the stated goals.

What is STEP?

STEP is an acronym for Standardized Test of English Proficiency. It is designed in an objective and unbiased manner to test a person's level of proficiency in the use of English. The test is made up of the following four components:

1. Reading Comprehension (RC – 40%),
2. Structure (ST – 30%),
3. Listening Comprehension (LC – 20%),
4. Compositional Analysis (CA – 10%).

The Committee chose these four components and their respective weights after consultation and research into the approach of both regional and international bodies responsible for the production of similar English tests and feedback provided by national organizations requesting such a test. Contingent on statistical analysis and feedback on STEP results, the weighting of the four components may be subject to revision depending on the needs of the beneficiaries. It should be added that STEP is uniquely designed by English language professors and experts native to the region and therefore knowledgeable of the abilities and the needs locally. This drastically lowers the possibility of cultural and social bias and provides a more appropriate tool for objective assessment.

Who should take the test?

STEP should be used for those who are seeking:

- admission at a university or college as an undergraduate.
- admission to a graduate program.
- employment in the private sector.

How STEP is developed?

STEP is developed carefully through several stages as demanded by best international practice.

▪ Selecting Item Writers:

The Center selects item writers either directly or by way of its contact persons in academic and educational institutions; item writers must be nominated by persons who have direct contact with the Center and who know the code of ethics the Center goes by.

▪ Training of Item Writers:

The Center will then hold a four-day workshop for item writers consisting of an intensive STEP orientation on how to write items of the multiple choice type; several intensive sessions are given on item analysis, test reliability and test validity. In the last two days of the workshop, item writers are asked to write items. For this purpose, they are divided into groups for hands-on experience in item writing. Each group is joined by experienced item writers and specialists in measurement.

▪ Actual writing of items:

At the end of the workshop each participant is asked to write between 100 and 150 actual items. Each and every item submitted is prepared for professional review by the Department of Testing and Measurement. Each item writer and every item submitted are assigned a coded number.

▪ Reviewing Committees:

With anonymity of the item writer assured, each item is reviewed by a committee which consists of three members: (1) a field specialist, (2) a specialist in measurement and (3) an expert who may be an advocate for the test-taker and the society at large. The Reviewing Committee scrutinizes each item for the following: content, domain, bias towards sex or region, difficulty, creativity, suitability of each item's stem and choices.

▪ Reviewing Committee decisions:

The Reviewing Committee then decides whether to accept the item as submitted, modify it, or reject it. This decision is based on charts constructed by the Department of Testing and Measurement and on their verbal instructions and reminders. All items are then stored in an item bank.

▪ Quality Assurance:

In addition to following the scientific steps in the construction of STEP, the quality is likewise insured by asking native English speakers to:

1. review the test at various stages of its formation;
2. revise and record the listening section.

Test Formation

The actual STEP test has 100 items distributed among the four components previously mentioned. With additional time allotted for non-scorable trial items and instructions, the total test time is 2.5 hours.

Statistical Criteria

Items are evaluated statistically with both Classical Test Theory (CTT) and Item Response Theory (IRT) models. Items are included if their CTT difficulty is reasonable (average difficult for the total test is around 0.5) and their discrimination is around 0.3. They also have to meet the criteria of the IRT three-parameter model:

- good discrimination,
- coverage of the ability spectrum,
- low guessing parameter value.

STEP Reliability and Validity

STEP reliability for the first version was 0.93, which indicates that the STEP is highly reliable. Initial validity was calculated for the relationship between the STEP score and the High School English Test of the Ministry of Education. The findings indicate that the STEP Standardized Test of English Proficiency correlation is significant ($r = 0.47$). Furthermore, STEP correlated at .68 with IELTS. and at .76 with the TOEFL, indicating a strong relation with those tests. It also differentiated between student levels in the English departments at the university level.