



Multilateral Comenius project awarded with European Language Label in 2009 and 2010



Alexandru Mîță

Paula Martins

Przemyslaw Stolarsky

Mihaela Liteanu

Roxana Elena Irina

Anna Maria Montuori

In-Service Training for Language Teachers

1 -Theoretical Issues

<http://www.lang-platform.eu>

The Londoner Publishing House 2011 – ISBN 978-973-88480-6-1



DG Educație și cultură

Programul Învățare pe toată
durata vieții

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Testing Writing

The Writing Tests have been undergoing changes from the past to the present. From the ancient essay tests to the latest TOEFL New tests and IELTS Examinations there have been aspects which have remained the same as fundamental principles and task types and there are other evaluation procedures that have undergone remarkable changes.

Essay tests are first analysed in detail and their drawbacks are pointed out. This is followed by an analysis of the interlinear tests, the objective tests and the pragmatic tests and a comparative study of these tests helps in drawing implications for testing writing. Finally, a review of a number of significant studies of writing research from 1900 to 1986 is undertaken and the overall implications are worked out.

This section begins with an overview of the historical perspective of written examinations and then goes on to discuss different types of tests like the essay test, the interlinear test, the objective test and the pragmatic test. A comparison of the different types of tests helps in drawing implications for the testing of writing.

Historical Perspective

Written examinations have a long history. Kuo reports that the Great shun, a model emperor of antiquity whose reign ended in 2205 BC tested his officers through written examinations every third year and after three examinations either gave them a promotion or dismissed them from service. European universities from the beginning emphasised examinations as the basis for the awarding of degrees and honours. The Cambridge Tripos is a well-known example of such examinations. In the 19th century Horace Mann administered uniform written examinations to a selected sample of learners from the Boston public schools. His arguments were influential in bringing about substitution of written for oral examinations. Horace Mann (1845) advocated the use of a large number of specific questions in place of fewer general questions and a search for more objective standards of educational achievement.

This search for objective evaluation of written work continues even today. While the teaching of writing emphasise the need for developing individuality, creativity and personal involvement, the testing of writing looks for the common factors which can be objectively evaluated so as to measure the learners' proficiency in written expression.

Essay Tests

Educators have realised the value of extended discourse for assessing learners' understanding and for interpreting their academic and personal experience, since the time of Aristotle. Essay examinations are widely used in school systems throughout the world. They are given during school terms to monitor the development of subject matter and writing skill, at the end of the courses and schooling levels to certify achievement, and as entrance examinations to determine qualifications for admissions to a higher level.

Though many countries have now abandoned essay examinations in favour of more easily scored objective tests, in India we still use it to a large extent. Even in countries like the US where multiple choice testing has replaced essay examinations at all levels a lot of criticism is voiced. It is said that the learners' ability to engage in disciplined thought and the ability to express it in coherent, supported discourse is seriously deficient. The NAEP (National Assessment of Educational Progress, 1981) challenges the construct validity of multiple choice questions for measuring the subject matter and especially writing skill development. They think more emphasis should be placed on test formats

which stimulate learners to use higher level reasoning processes as they formulate extended discourse, solve complex problems or apply subject matter concepts and principles. The ETS (Educational Testing Service, 1983) research report describes the attempt made to determine the extent to which the current version of TOEFL is a valid indicator of the English Writing Skills' which mainly consists of multiple choice items. TOEFL (Test of English as Foreign Language) has conducted a survey research to suggest that a well-constructed test can serve both the functions. The same test could help us evaluate the level and development of the learners as well as compare them normatively to the progress of other population of learners.

a. Components of Essay Examinations

Essay examinations have often been criticised for their lack of objectivity. Critical components of essay examinations methodology need to be identified and analysed, if we want to make better use of essay examinations.

Some of the important components identified are as follows:

- The skill to be assessed,
- The essay problem assignment, and
- The scoring criteria

The skill to be measured should be sufficiently defined if the test is to be reliable. The first step in formulation test specifications is to define the skills assessed. Skill specifications are generally considered to include definitions of the content and behaviour. In subject matter essay examinations the content would be facts, concepts and principles. The behaviour would not be just the observable behaviour "write", but the procedure, strategies and solution routines the learner was to apply to the content. In tests of writing ability the focus is not on content points covered but on the discourses features of the requested essay, for example, whether the composition is a well-formed example of narrative or expository writing.

b. Structures of assignment

In tests of writing ability the structures of assignments have varied considerably. At one end of the continuum essay tasks are described as topics. Topics could be a simple one word clue which might bring out different kinds of responses. At the other end of the continuum essay tasks direct learners treatment of discourse, mode, topic and audience or the reader. The one word prompt implies the view that writing is a skill that can be equally demonstrated in response to any one of the myriad writing tasks. The kind of writing tasks presented in class is often criticised as they do not present full rhetorical contexts that sufficiently inform the learners about the writing purpose, topic, audience, writers role and intended criteria for judging the essay (Britton1975).

Recent rhetorical studies suggest that different rhetorical purposes (to express, persuade, inform) place different cognitive demands on the writer and consequently, learners write differently when writing for different rhetorical aims and audience. In a study conducted by Quellmalz, Capell and Chou (1982) writing competency profiles derived by tests differing in discourse mode and response mode were compared. Their findings establish the fact that levels of performances vary on tasks with different writing purposes. Measures of writing which demand different discourse modes, such as narrative or expository, tap different cognitive skills and hence it is important to clarify the demands that are required by different, specific writing tasks.

Tests of written assignments should also take into account the time allotted for planning, writing and at least reviewing, if not revising the written work. The amount of time scheduled for essay writing varies widely across countries. Time can support or constrain learners chance to demonstrate their competence. The current test theory and research support the advisability of structuring essay

examination prompts that clearly specify the aim, topic, audience, writers, role and evaluation criteria and that which permit sufficient time for learners to engage in all aspects of the writing process.

Scoring criteria & Rating scale formats

The criteria used to judge the essay examination operationally define which content features and test structure constitute a “good” or at least a “competent” response. To be credible, criteria should not reflect the preferences of only a few individuals, but should represent standards endorsed by a community of professionals knowledgeable about the subject matter.

Secondly the criteria should refer to these features of content and written expression, which are amendable to instructional intervention. We cannot test what we do not teach in the classroom. For example, dimensions of “depth”, “flavour”, and “creativity” may enhance the quality of the essay but a growing number of educators contend that it is neither logical nor fair to hold the learners accountable for subject matter or writing expertise that the schools cannot demonstrate they can teach.

The criteria used to evaluate learners’ content and written expression vary along a number of dimensions. The variation may be as follows:

- From qualitative value judgements to quantitative counts of information and test features;
- From global reactions to analytical judgements;
- From comprehensive attention to a range of concepts and text features to isolated focus on particular information or text feature;
- From vague guidelines to replicable precise definitions.

Generally, readers’ reactions to learners’ essays involve three levels of judgement.

- 1) Subjective, global impressions of overall quality
- 2) Analytic judgements about component test features
- 3) A holistic quality judgement combining subjective impressions with judgements about the quality of the combinations of text elements.

Global judgment

In general impression scoring, a rater reads an essay once and assigns it a quality score. General impression ratings are global, heavily qualitative and are based upon vague guidelines that may not refer to component text features or their differential weighting or importance.

Analytic judgement

The most quantitative, detailed and replicable scales are analytic rating scales where readers assign several scores for various features of the essay. Analytic scores vary considerably in the range of content, rhetorical, structural and syntactic elements referenced and in the relative weights of these elements. The analytic scores differ in the importance they give to different features of written assignment.

S Mohanraj (1981) discusses analytical rating scales of Carroll (1961), Alan & Campbell (1965), Cooper (1972), Davies (1977) and Pilliner. He has prepared a model of his own which includes twelve features of writing. He has further simplified it and has arrived at a model suited to our situation where teachers cannot spend much time in correcting compositions. This model is quite practicable and easy to use.

A similar model is suggested by suggested by Rita M. Deyoe (1980). Her model gives more importance to grammatical aspects whereas Mohanraj's model attempts to concentrate on stylistic and discursal features.

Holistic judgement

Holistic scales, where readers assign a single score, often combine characteristics of both general impression and analytic approaches. Holistic schemes vary widely in the range of text elements contributing to each score point and the specificity with which score levels are defined (Ingenkamp 1977, Quellmalz 1980).

Since the focus, specificity and objectivity of criteria informing impressionistic, holistic or analytic approaches vary considerably, an examination programme should weigh carefully the nature of the criteria selected and their underlying rationale. Otherwise the programme may find that the criteria do not match well with the aims of the assessment and instructional programmes and do not provide a useful status report or diagnostic feedback. The need for explicit criteria is also apparent for scoring subject matter essay examinations. Learners commonly complain about the ambiguous subjective criteria used for subject matter essay examinations in the classroom assessments. When results of large scale achievement exams have serious consequences for learners' explicit public and rational scoring keys are imperative.

Rating Procedures

When a large number of papers must be scored by a pool of readers, an assessment programme must ensure that evaluation criteria are uniformly interpreted and applied. Such standardization involves both the formulation of explicit criteria and procedures for training raters. In the US rater training follows a fairly standard procedure. The following steps are employed to train raters.

There is a brief introduction to the rating scale.

Then the raters begin to practice applying criteria to a set of papers representing the test sample.

A trainer leads a discussion of the features of each paper that result in the classification of the paper to a particular grade.

Training time varies according to the number of separate scores recorded for each paper and according to the clarity of the criteria. The rigor of the procedures used to decide if acceptable rater agreements levels have been attained at the end of the training vary from a show of hands to pilot tests requiring independent scoring of essays.

In India through essay examinations are widely used, there is no programme to train raters. Failure to conduct any structured training or to check on prior agreement levels may increase the risk of unreliable scoring.

Reliability

The reliability of an examination programme depends on the degree to which it eliminates measurement error. Four potential sources of error or score fluctuations identified for examinations of writing ability (but applying as well to tests of subjects matter skills) are as follows:

The writer – within – subject individual differences.

The assignment variations in item or task content.

Between - rater fluctuations

Within - rater instability

The writer within – subject errors can be avoided if the learners are asked to write a series of essays instead of one single essay. Thus the reliability of learners' performance can be determined by

gathering data on a pool of homogeneous items or assignments. Since essay writing requires at least twenty or thirty minutes it is often difficult to have them write many essays in examinations. But studies of the consistency of learners' performances across a series of essay often report low reliabilities for a single essay. According to Spencer (1979) analysis of the stability of learners writing performance across several essays is also not reliable because of the variability brought in by the difference in topics.

Some ways of overcoming the problem of reliability are as follows:

- Essay tasks should be based on specific skills of writing. This would reduce error variance due to the assignment.
- Essays should be collected on at least two parallel assignments. This would reduce error associated with individual variability.
- Scores on several essays should be combined to increase the readability of subject matter essay examinations.

Inter-rater agreement is the most prevalent issue concerning reliability in essay examinations. Statistical indices of agreement levels include co-efficient alpha, generalisability co-efficient, point biserial correlations and simple percentages of agreement. The most effective method of reducing inter-rater variability is to provide training on clearly specified criteria. To reduce error due to within – rater score fluctuations over time (rater drift) due to reader fatigue and / or carelessness, some form of interspersed check procedure seems helpful, according to Quellmalz (1980). Although some studies report that readers tend to get more lenient or harsher as rating progresses, few assessment programmes routinely monitor this problem.

Mike Hayhoe (1983) in his article, 'A Historical Review of Essay Marking' discusses the problem of reliability in marking essays. According to him this problem has been persistent for a long time in the history of marking essays. If Rowntree was concerned about marker reliability in the 1880's, Raleigh (1980) is equally worried about the same problem. Mike Hayhoe says that an error of twenty five percent in grading an essay may be conservative estimate and it has been suggested that the problem of unreliability in marking essays exists in internal assessment as well as external.

Reliability is inextricably linked with validity. The reliability of an essay examination depends on how valid the examination is and how valid the markers are in their assessment. A brief consideration of the problems faced by examiners in designing valid examinations is necessary if one wants to integrate testing and instruction.

Validity

The validity of an examination derives from evidence that the test accurately and dependably measure the specified skills. Evidence for the validity of an examination may take several forms.

One form focuses on the test content, that is, the test items or essay assignment, and gathers judgement of subject matter experts regarding a number of things like -

- the objectives or skills defined to be important and representative of subject matter competencies, and
- The way these skills are elicited in the item, problem or writing assignments.

Other forms of validity focus on test performance to examine the following things:

- Concurrent validity – whether the scores are comparable to scores on other tests of the same skills,
- Predictive validity – if the score levels predict future success, and
- Construct validity – if the performance pattern appears to measure the underlying trait.

The most common methods of attempting to establish the validity of essay examinations have been comparisons of scores to 'related' measures. In the case of tests of writing ability, the 'other' measures chosen as criterion variables are often reading tests, multiple choice or class grades.

The heart of the validity of a test is whether it measures the underlying skill construct, that is, whether it taps the hypothetical mental store of information and strategies. According to Raleigh (1980) the validity of an examination can be described in terms of the degree to which it 'measures well' what it is intended to measure. According to Mike Hayhoe (1983) there is a possibility to think of 'Markers' validity that is the degree to which he 'measures well' what the assessment systems sets out to measure.

Factors affecting marking

The marks awarded to an essay depend on a number of things. For example, Thorndike (1986) discusses the problem of 'uniqueness'. Uniqueness raises the issue of divergence, the individuality of the work, and convergence, notions of correctness and orderliness. How far a marker is affected by divergence and convergence will decide the marks he gives to a particular assignment. Wiseman and Wrigley (1958) identified two schools of thought as far as assessors' value base are concerned. One school values 'imponderables' of validity, freshness and fluency. The second school of thought sees the writer as 'a craftsman able to show his skill whatever type of materials he works in'.

Britton (1963) found some evidence to suggest that teachers may well group towards valuing one end or the other of the following two poles:

Sophisticated, conventional written based work

Work based on familiar speech

Work based on imagination including fantasy / the unreal

Work based on observation of real life

A number of studies conducted in America suggest that teachers tend to cluster in favouring certain criteria ideas, form, flavour, mechanic, wording - and that the clusters of criteria adopted by the teacher can affect grading.

Deale (1975) feels that 'adequacy' of writing rather than ideas affects the marks awarded. Soloff (1973) argues that lack of consonance between the writers' values and those of the assessor on a topic may affect the grade awarded. The London Association for the teaching of English shares his opinion. In its pamphlet, *Assessing Compositions* (1965) it expresses concern about how an assessor may react to experiences and attitudes in an essay which are unfamiliar to him and the potential for under or over assessing the work.

Marshall (1960) suggests that assessment in terms of the features of pieces of work which 'float' to the examiner - his intuitions about the texts - is the proper activity of an alert and sensitive marker.

Markers can be affected by visual features at the expense of such aspects as organisation, fluency, appropriateness in terms of task, audience and so on. According to Mike Hayhoe (1983) this may be because the visual features are more immediately obvious, especially when they are flawed, and because there is a greater degree of consensus about them than there is about what 'coherence' or 'clarity' or other more global criteria may be.

Marshall (1967) and Scannel (1966) have found assessors particularly adversely affected by spelling errors, with errors of grammar and punctuation coming next. Handwriting also has a great impact on the assessors and many researchers like Chase (1968), Briggs (1970), and Soloff (1973) have

demonstrated the power of this feature in affecting marking. In his more recent work Briggs (1980) goes further, suggesting that there may be borderline areas in grading in which this value aspect of a piece of writing may be the major factor in deciding what it is worth.

Yates and Pidgeon (1957) found that the setting of an essay affected the markers' response. If an 'average' piece of work followed several fine pieces, it was likely to be marked hard; if it followed several poor ones it was likely to be upgraded.

The analysis of the present situation in Gujarat also reveals the fact that teachers are more concerned with spelling errors and punctuation. Next comes the grammatical error. Though all the teachers marked a number of features in the questionnaire (appropriacy, organisation, overall writing ability etc.) as very important, all of them assign one single grade on the basis of the overall impression of the composition.

Drawbacks of Essays Examinations

Essay examinations are said to test learners' ability to engage in disciplined thought and the ability to express it in a coherent, supported discourse. But a number of points need to be taken into account if essay examinations are used to measure writing ability.

Some of the problems involved in using essay type tests are as follows:

It is difficult for an average teacher to structure such prompts for essay tests that clearly specify the aim, topic, audience, writer's role and evaluation criteria. The problems of reliability, validity and the factors that affect marking discussed in this section prove that it is very difficult to measure the skills of writing ability through essay examinations.

Teachers cannot spend a lot of time in checking essays using analytic or holistic rating scales. The general impression score usually assigned by teachers is not a reliable method of scoring.

The method of training of raters is expensive and time-consuming and is not practicable as far as the schoolteachers are concerned.

Since it is not easy to structure, administer and score essay examinations, we need to consider other types of tests which are easy to construct, are easy to evaluate and which give a reliable and valid indication of learners' proficiency to communicate through writing.

The interlinear test

Findley and Warren (1953) made an attempt to overcome the defects of essay examinations arising from its unreliability of scoring, by the development of 'semi – objective' tests of writing ability termed the Interlinear Test. This test overcomes most of the objections to objective tests of writing ability. Objective tests usually present the learner with poorly written material and provide him with several options regarding the corrections, which should be made. According to Swinford, France (1956), the interlinear test structures the learner's response to such an extent that the reliability of scoring is of a very high order.

The interlinear test presents the candidate with a triple spaced copy of a badly garbled piece of writing. The learner is allowed thirty minutes to indicate the necessary co-relations and deletions but is instructed not to add his own ideas. The learner's paper is then 'scored' for his treatment of predetermined errors. The reliability of the interlinear test compares favourably with that obtained for

objective tests of the same working time and length. The validity of the interlinear test also is as valid as that of the objective test.

It would be useful to analyse the objective test in detail so as to get a clear understanding of its advantages and disadvantages.

The Objective Multiple Choice Test

The objective type test may be of the short answer type or the choice item type. A number of studies have considered the relative validities of short answers and choice items. Cook (1955) found that the correlation between scores of knowledge of contemporary affairs for college students and multiple-choice test over the same item of information was so high, when correlated for attenuation, that it was impossible to say that the two tests were measuring two different kinds of achievement. Hurlburt (1954) on the other hand, reported significant differences between completion and multiple-choice tests as measures of precise knowledge of word meaning.

Short answer questions

The short answer items require the learner to write a word, phrase, number or symbol, while the choice items ask him to choose one of the many alternatives given in the test item itself. The short answer type questions may be of the recall of a word or number. The choice items often present new problem situations and suggest alternatives, which the examinee has never seen before. So they call for more than recognition.

Short answer questions make a heavy demand on the test constructor because of two reasons:

- He must concentrate on detailed and highly specific questions to which brief but quite unique verbal or numerical answers can be given.
- He must be prepared to deal with frequent and difficult decisions as to whether a given answer is just barely adequate or not quite adequate.

Choice items may be one of the following types:

- Classification
- Matching
- Multiple-choice
- True / false
- Rearrangement items.

The advantages of the choice items are as follows:

Choice items can be made very difficult, without involving ambiguity and as searching of high levels of competence as items in any other form.

To make wise choices it is necessary to perceive clearly the implications or consequences of choosing each alternative.

Multiple choice items

A number of books discuss the construction and administration of multiple choice items. 'Language Testing' by Robert Lado (1961), 'Modern Language Testing' : 'A Handbook' by Rebecca Valette (1967), 'Testing English as a Second Language' by David Harris (1969), Foreign Language Testing : Theory and Practice by John Clark (1972), 'Testing and Experimental Methods' by J P B Allen and Alan Davies (1977), Revision of Modern Language Testing by Valette (1977) are some of the important books which deal with the construction of discrete point tests.

Discrete point multiple-choice tests assess one skill at a time, listening, speaking, reading or writing. They assess only one aspect of the skill – i.e. productive versus receptive, oral versus visual, etc. They attempt to focus attention on one point of grammar at a time. Each test item is aimed at one element of a particular component of a grammar item. According to Lado (1961) within each skill, aspect and component, discrete items focus on precisely one and only one phoneme, morpheme, lexical item, grammatical rule or whatever the appropriate element may be.

Remmers and Adkins (1942) studied the reliability of multiple-choice tests as a function of the number of responses per item. They found that reduction in the number of distracters tended to lower the test reliability. Spearman - Brown formula gave reasonable good predictions of the reduced reliability when distracters were eliminated at random. Williams and Ebel (1957) started with four-response forms by systematically eliminating the least effective distracter. They found that in a test period of fixed time limit, a greater number of two response items would produce more reliable scores than a smaller number of three or four response items. According to Weitzman and Ellis (1946) the essential characteristics of the distracters of multiple-choice items is that they should be plausible to those who lack the knowledge or ability for which the item is testing. Hence a lot of care should be put into the selection of the distracters.

Drawbacks of discrete item tests

The construction of multiple-choice items is very difficult. Lot of care needs to be taken while preparing the 'stem' or the main sentence and the different choices from which the learner has to select one. The main sentence or question should be unambiguous and explicit. All the choices should look plausible for those who lack the knowledge. At the same time there should not be more than one correct answer.

Standardizing discrete item tests require pre-testing, statistical evaluation and rewriting techniques, which is quite time-consuming and require a lot of technical knowledge. Hence it is not practical for an ordinary classroom teacher who may not have the time or technical skill to prepare these discrete item multiple-choice tests.

John W Oller Jr. (1979) criticizes the discrete item tests. According to him discrete point tests view language as form and usage rather than of process and use. In actual language use, language is not used in bits and pieces. A proficient user of language is not overtly conscious of the isolated phonemes or graphemes or any such bits of language. Again proficiency in using isolated bits does not indicate that the learner can put together these bits and use the language for communication. To test language as communication we need to think of a different kind of test, which tests the learners' ability to use the language.

Types of Language test Based on the Criteria of Administration

1. Selection test

- Chose the candidate that has fulfilled in the certain condition.

- The candidate will be accepted or rejected.

2. Placement test

The test just intended to place who tests takes

- Commonly, it's giving before the program is begun
- The objective is to take place the candidate based on the group or the level of their ability.
- It's giving if the candidates too much to be in one group.

3. Achievement test

Is a test which is giving to know how much the material that the students' have already achieved after joining the learning process/instruction.

4. proficiency test

Proficiency test is a test to know the ability or proficiency of the participants. Proficiency tests are designed to measure people's ability in language, regardless of any training they may have had in that language.

5. Diagnostic test

- The objective is to know the strength and weaknesses of the students.
- To know the difficulty study of the students, so the teacher can decide the material or planning that will suppose to do.

6. Try out

Is a test that normally conducted before the real test is doing.

- To know weather the real test that giving is suitable with the time allocation option, instruction, etc.). And have a good characteristic such as.
 - a. Valid
 - b. Reliable
 - c. Practicality

Types of Language Test Based on the Criteria of Admission-Time/Steps

1. Admission test

Admission test (entrance test) is the test that held before a program is begun. The objective of the test is to decide whether someone will be accepted or rejected. This test is also used as selecting test. For example: As a guide

2. Formative test

Formative test is the test that held to know whether or not the program is run a way. It gives to the students or participants in order to plan what is should to do after doing the test.

3. Summative test

Summative test is the attempt to summarize students' learning at some point in time, in the end of the course.

4. Pre-test

Pre-test is preliminary test administered to determine a students' baseline knowledge or preparedness for an educational experience or course of the study.

5. Post-test

Post-test is a test given after a lesson or a period of instruction to determine what the students have learned.

Types of Language Test Based on the Criteria of Doing/Answering the Test

1. Written-test

Written-test is tests that usually do in written. For example, Reading comprehension test and Writing test.

2. Spoken-test

Spoken test is tests that usually do orally. It usually uses to measure the speaking ability of the students'. For example: testing speaking.

Types of Language Test Based on the number of candidate

1. Individual Test

The test given to a candidate each candidate gets its own test. This kind of test intent to measure language skill that needs to measure effectively. So that usually this kind of test applied in speaking test which needs to measure it achievement directly. Ex: interview

2. Group test

The test administered for a certain candidates. The aim of this kind of test is to make the test more efficient in term of time and energy but not only that this kind of test is also good to test speaking ability in term of interaction.

Types of Language Test Based on the way of answering

1. Essay test

Test that requires candidates to answer questions in writing. Responses can be brief or extensive. Test for recall, ability to apply knowledge of a subject to questions about the subject, rather than ability to choose the least incorrect answer from a menu of options.

2. Short answer

In a short answer question, the candidate types in a word or phrase in response to a question. The answer could be a word or a phrase, but it must match one of your acceptable answers exactly.

3. Multiple choices

Test in which students are presented with a question or an incomplete sentences or idea. The students are expected to choose the correct or best answer/completion from a menu of alternatives, Option that consist of a key and destructors.

Types of Language Test Based on the way of scoring

1. Subjective Test

Test in which the impression or opinion of the scorer determine the score or evaluation of performance.

2. Objective test

A test for which the scoring procedure is completely specified enabling agreement among different scorers. It means that we can make answer key in advance and the correct answer must be the same with the one in answer key.

Types of Language Test Criterion of Reference Score

1. Norm reference test.

In held the Norm reference test, interpret to change the rough score become the final score, it can be done based on the level of the average from one group of the candidates that doing the test.

2. Criterion reference test.

In the Criterion reference test, the result of the final test doesn't has correlate with the result from the another candidates that do the same test. The final score on the criterion reference test, based on the result of low languages capability level that can be accepted as the standard of language capability.

3. Mixed both of Norm reference test and Criterion reference test.

In applying on daily test, result of the final score can't based only one from two references. There are the final score built by combine both of them.

Types of Language Test Criterion Aspects of Language

1. Languages Talent tests.

Languages talent test it means to know and measure the talent and ability potentially that someone has for learn languages.

2. Languages Skill tests.

With languages skill test, we can get information about level of ability by using languages to the process of something. Information that had from the languages skill test it's about level of real ability at that time. Example languages skill test are reading skill test, speaking skill test, and writing skill test.

3. Languages Component test.

Languages can be dividing in many components and each component is different. The components are pronunciation, vocabulary and grammatical. In structural approach, teach languages means teach how to mastering the components of languages. Based on this approach, language component can be measured with pronunciation test, vocabulary test and grammatical test.