

## LANGUAGE TESTING: CHARACTERISTIC OF GOOD TESTS, TESTING LANGUAGE SKILLS AND COMPONENTS

Tiray Febria Zananda, S.Pd., M.Pd<sup>1</sup>, M. Lucky Setiawan<sup>2</sup>  
Palembang University<sup>1</sup>  
febriatiray@gmail.com

### ARTICLE INFO

#### Article History

Submission: November 17, 2023  
Review: November 19, 2023  
Revised: November 19, 2023  
Accepted: November 19, 2023  
Published: November 19, 2023

#### Keywords

Ability  
Language testing

### ABSTRACT

Tests can give contributions if it is use naturally. It is needed to do a test to know students' ability. Tests are also used to measure ability in language. Language testing is a field of study under the umbrella of applied linguistics. The purpose of testing is to collect objective information that may be used in conjunction with subjective information to make better. Language test developers and teachers should think of new types of intensive tests to measure examinees' real language competence and performance to make the tests more reliable.

### ARTICLE INFO

#### Article History

Submission: November 17, 2023  
Review: November 19, 2023  
Revised: November 19, 2023  
Accepted: November 19, 2023  
Published: November 19, 2023

#### Keywords

Kemampuan  
Tes Bahasa

### ABSTRAK

Tes dapat memberikan kontribusi jika digunakan secara alami. Perlu dilakukan tes untuk mengetahui kemampuan siswa. Tes juga digunakan untuk mengukur kemampuan berbahasa. Pengujian bahasa adalah bidang studi di bawah payung linguistik terapan. Tujuan pengujian adalah untuk mengumpulkan informasi obyektif yang dapat digunakan bersama dengan informasi subyektif untuk menjadikannya lebih baik. Pengembang tes bahasa dan guru harus memikirkan jenis tes intensif baru untuk mengukur kompetensi dan kinerja bahasa peserta ujian yang sebenarnya agar tes tersebut lebih dapat diandalkan.

### Introduction

Tests may be constructed primarily as devices to reinforce learning and to motivate the student or primarily as a means of assessing the student's performance in the language (Heaton, 1988:5). Tests can give contributions if it is use naturally. It is needed to do a test to know students' ability. Tests are also used to measure ability in language. Language testing is a field of study under the umbrella of applied linguistics. Its main focus is the

assessment of first, second or other language in the school, college, or university context; assessment of language use in the workplace; and assessment of language in the immigration, citizenship, and asylum contexts (Wikipedia, 2013). In language testing, it may include listening, speaking, reading, writing or cultural understanding. Moreover, it also may be placed on language theoretically and proficiency. In addition, Oller (1979: ) explains language testing is a device that tries to assess how much has been learned in a foreign language course, or some part of a course. This paper will explain more about characteristics of good tests, testing language skills and component.

## Literature Review

### Language Testing

Language testing is a device that tries to assess how much has been learned in a foreign language course, or some part of a course (Oller, 1979 :). Fulcher and Davidson in Nguyen & Le, 2012) adds language assessment or testing is an important stage in the process of language teaching and learning as it helps monitor students' educational progress and evaluate the quality of school systems. It can be conclude that language testing is a evaluation process that needed to measure students' ability in learning language.

Language tests examines as:

- a. Tests
- b. Investigation of learner characteristics by using language tests as elicitation procedures.
- c. Specific hypotheses about psycholinguistic and sociolinguistic factors in the performance of language based tasks may be investigated using language tests as research tools.

Testing should create a positive wash-back effect, i.e., the test should positively affect teaching. This is achieved by making sure that the test actually tests what has been taught, and does not omit to test anything important. For example, if a language test does not have an oral component, teachers will probably not test pronunciation and learners will not see pronunciation as important. Do not use any question type (true/false, multiple choice, comprehension questions, essay) that is not familiar to the candidates. Make sure that any objectively scored questions (T/F, MCQ) are 100% unambiguous. This is usually best done by having two or three colleagues check the test before it is administered. Make sure subjectively scored questions (essays) have a clear set of marking criteria: 'Candidates will demonstrate knowledge of / ability to do A, B and C' so that all markers know what every answer must contain. Set a reasonable word limit for the time available to candidates. If you use multiple choice, all the choices for a question should be of the same length and with similar phrasing, as candidates will often go for the option that stands out from the others

rather than the correct one. If you use comprehension questions about a text, make sure candidates cannot answer them simply by quoting directly from the text. Exam papers should have ‘face validity’, i.e., the papers should look professionally produced, be easy to read, no smudges or typing errors, have clear instructions and an example for each section. Otherwise candidates are faced with problems not connected with the subject being tested.

The purpose of testing is to collect objective information that may be used in conjunction with subjective information to make better educational decisions Kubiszyn, T. & Borich, G. (1993). The remarkable contributions the language testing are help create positive attitudes towards instruction by giving students a sense of accomplishment; help students learn the language; provide a measure of students’ achievement in and upon the completion of language instruction; provide diagnostic tips as feedback for teachers; define course objectives (Madsen, Bowen et al, Valette in Burhan, 2009:72). Therefore, from the tests, it can make students become more motivated in learning because of their achievement. Teacher also sometimes can give the test easier so students can answer it to make them more motivated in learning. Tests also can improve the program; it can be seen from students’ score.

In contrast, Gipps (1994, 34-37) explains the impact of testing on the curriculum and teaching. The power of tests and exams to affect individuals, institutions, curriculum or instruction is a perceptual phenomenon. Actually, the thing that important in testing is not about the high or low score, but the most important is students’ effort in doing the test and the impact of the tests itself in their learning; The more any quantitative social indicator is used for social decision making, the more likely it will be to distort and corrupt the social processes it is intended to monitor. The social indicator can be change both substantial and corrupting; If important decisions are presumed to be related to the test results, then the teachers will teach to the test. The evidence seems that to make score of the test rise, the skill itself should improve so as a teacher should make their students pass their test because as their professional duty; in every setting where a high-stakes operates, a tradition of past exams develops, which eventually de facto defines the curriculum. To make a good exam result, school should have good curriculum, it means strategy that teacher use to drill their students. Teachers pay particular attention to the form of the questions on a high-stakes test and adjust their instruction accordingly. For example, short answer essay, multiple choices; when test results are the sole, society tends to treat test results as the major goal of schooling, rather than as a useful but fallible indicator of achievement. Actually, people think only the result rather than what is the benefit that they get at school (indicator of achievement); a high-stakes test transfer’s control over the curriculum to the agency which sets or controls the exam.

## Characteristics of Good Tests

Good tests should include these four point: valid, reliable, practical, and instructional value.

### A. Validity

- a. Related to how well the test does what it is supposed to do
- b. Inform us about examinee's progress toward some goal in a curriculum or course of study
- c. To differentiate levels of ability among various examinees on some task
- d. Question: What a test actually measures in relation to what it is supposed to measure.

A test is said to be valid if it measures what it claims to measure. Validity is a unified concept. As such, it has several dimensions or aspects, they are:

#### a. Face Validity

It relates to how a test looks to other people, students, experts, etc.

Example: a grammar test should test the grammar not the vocabulary. Thus, in a grammar test, the vocabulary should be easy and vice versa.

#### b. Content Validity

A test needs to have a representative sample of the teaching/instructional contents as defined and covered in the curriculum. How do we know that a test is valid in content? By using blueprint/kisi-kisi/rambu-rambu.

Example: a grammar test should contain every part of the grammar. But we cannot take all of them, so just take the sample to represent the content.

#### c. Empirical Validity

It relates to the closeness between the score obtained from a test with the other criteria outside that test. It is divided into two:

- a) Concurrent validity: how well the test estimates current performance on some valued measure other than the test itself.

e.g. TOEFL and TOEIC >> if someone's TOEFL score is high, then we can say that automatically her/his TOEIC score is high too.

b) Predictive Validity: how well the test predicts future performance on some valued measure other than the test itself.

e.g. GPA >> if someone's GPA is high, we can say that s/he will have a high salary job.

#### d. Construct Validity

It relates to our understanding of the existing theory to construct a test.

Example: if we want to construct a speaking test, we have to find, compare, and criticize the theory about the speaking test not the other.

Construct validity requires two levels:

##### 1. Logical Analysis

ü Dimension >> standard of competence

ü Variable >> basic competences

ü Sub variable

ü Indicator

ü Test items

##### 1. Empirical Analysis

Whether the test items measure what is defined by their indicators.

##### • Washback Validity

It relates to the influence of a test to the teaching learning process. There are two types of washback:

##### 1. Positive Washback

*Micro Level (classroom setting)*

ü Tests induce teachers to cover their subjects more thoroughly.

ü Tests make students work harder.

- ü Tests encourage positive teaching-learning processes.

*Macro Level (educational/societal system)*

Decision makers (govt.) use the authority power of high stakes testing to achieve the goals of teaching and learning, such as the introduction of new textbooks and new curricula.

1. Negative Wash back

*Micro Level (classroom setting)*

- ü Tests encourage teachers to make “teaching to the test” curriculum.
- ü Tests bring anxiety to both teachers and students and distort their performance.
- ü Tests bring anxiety to both teachers and students and distort their performance.
- ü Tests drag students to learn discrete points of knowledge that are tested.
- ü Tests make students create a negative judgment toward tests, and alter their learning motivation.

*Macro Level (educational/societal system)*

Decision makers overwhelmingly use tests to promote their political agendas and seize influence and control of educational systems.

Reliability

- a. Is a matter of how consistently it produces similar results on different occasions under similar circumstances.
- b. Question: how consistently a test does what it is supposed to do, and thus cannot be strictly separated from validity questions.
- c. Test cannot be any more valid than it is reliable.
- d. A test is considered reliable if it is taken again by the same students under the same circumstances and the score average is almost the constant, taking into consideration that the time between the test and the retest is of reasonable length.

Practicallity

- a. Must be determined in relation to the cost in terms of materials, time, and effort that is required.
- b. Including preparation, administration, scoring, and interpretation of the test.

#### Instructional value

- a. Pertains to how easily it can be fitted into an educational program, whether the latter involves teaching a foreign language, teaching language arts to native speakers, or verbally imparting subject matter in a monolingual or multilingual school setting.

### Testing Language Skills and Components

Language has been analysed as being composed of skills--listening, speaking, reading, writing and component-sound, vocabulary, structure . The language skills of listening contain the language components of vocabulary and structure; speaking--sound, vocabulary and structure; reading--vocabulary and structure as writing does. Language skills in English has four main major, they are listening, speaking, reading, and writing. According to Bowen et al in Burhan (2009:73), language tests are classified as skill tests and knowledge tests. Skill tests including listening, speaking, reading, writing while knowledge tests are those that test linguistic information, culture, literature, and the like apart from skills. There are two main types of language assessment: traditional assessments (paper-and-pencil language tests and performance tests) and alternative assessments (checklists, journals, logs, videotapes and audiotapes, self-evaluation, and teacher observations, etc.) (Brown & Hudson, & McManara, in Nguyen & Le, 2012).To do a test, especially language skills, in listening comprehension can use short utterances, dialogues, talks, and lectures are given to the testers; for speaking ability may use interview, picture description, role play, and problem-solving task involving pair or group work; In reading comprehension, it can be understand the gist of a text and to extract key information; Then, in writing ability may use letters, reports, memos, messages, instructions, and accounts of past events.

Language has been analyzed as being composed of skills--listening, speaking, reading, writing and component-sound, vocabulary, structure. The language skills of listening contain the language components of vocabulary and structure; speaking--sound, vocabulary and structure; reading--vocabulary and structure as writing does. Harris view was a static one as it ignored the dynamic nature of language use which causes the skills to be interchangeable requiring participants in the interactional and transactional discourses to negotiate ideas, feelings and understanding. Both skills and components are nuts and bolts of language. The former may be compared to bone or skeletal structure of a human flesh body while the latter to the human flesh. They must be together for form a complete living

human being. In language testing, to test language component alone void of skill is called discrete-point test.

### **The Evaluation**

When evaluating, the teacher plan an approach for evaluating the materials to determine what kind of changes need to be made in them. The evaluation is given in order to check wether the instructional is effective or not to continue. If it is effective, the instruction model can be continued to the next turn, can be with some additions or revisions. But if is not effective, the instructional model must be revised. Teachers can judge whether it is effective or not based on the students' test result.

The content will be accurate if the materials provided in the curriculum has been taught all and the students pass the standard grade or passing grade that is 75. If the students get score less than it, they must have re-treatment the materials and have remedial test. The result of the test will be analyzed by using certain form of analyzing whether the teaching and learning process is success or not.

#### **a. Formative test**

The output of each stage of design-goals, learner and context analysis, task analysis and so on-can all be submitted to formative evaluation in order to make revisions prior to any actual development of materials. In the teaching process, the formative test can be done four times until five times after the skills have been taught and the mid term test. During this phase of formative evaluation, the designer attempt to find out whether: (1). The instructional goal reflects a satisfactory response to the problems identified in the need assessment. (2). The environment and learner analyses accurately portray these entities. (3) The task analysis include all of the prerequisite skills and knowledge needed to perform the learning goal, and prerequisite nature of these skills and knowledge is accurately represented. (4). The test items and resultant test blueprints reflect reliable and valid measures of the instructional objectives. (5). The assessment instruments and their related mastery criteria reliably distinguish between competent and incompetent learners.

Goal review confirms that the goal that have been established are representative of a real instructional need and congruent with the client's expectation. The institution requires the ultimate goal of teaching English in the school is the students to be able to communicate English in daily life actively. Here it is same as what is the ultimate goal of teaching and learning English will be conducted. The designer will have gained much data as it developed from the analysis of the environment and the learners. The environment of the school is supporting enough for students to study English well, because they have English day on Friday when all schools persons parctice English and the school provides a native speaker. It will encourage students to practice English too. The task analysis can be confirmed using



a number of techniques taught, based on the skills. The prerequisite relationship of skills may be confirmed by testing groups of learners if the learners who can achieve the terminal objective can indeed perform all those listed as subordinate and vice versa. Assessment items can be affirmatively evaluated for their validity by having content and testing review the assessment item specification. In this case, teachers make questions based on SKL and the format of make test are included. So the teachers can evaluate the congruence of the objective and the item specifications and determine if the type of items outlined by the specifications adequately describe the domain of items that the objectives might cover.

During evaluation, the teacher tries out the instructional materials with one to one evaluation, small group evaluation, performance data, attitude data, time data and on going evaluation to the students.

## Conclusion

Language has been analysed as being composed of skills--listening, speaking, reading, writing and component-sound, vocabulary, structure . The language skills of listening contain the language components of vocabulary and structure; speaking--sound, vocabulary and structure; reading--vocabulary and structure as writing does. Language testing, as its specific purpose, is considered a gateway to education, employment, moving from one country to another and making decisions on the placement of students on particular courses (McNamara, in Nguyen & Le, 2012). It is important that language test developers and teachers should think of new types of intensive tests to measure examinees' real language competence and performance to make the tests more reliable.

## Declarations

I hereby declare that the disclosed information is correct and that no other situation of real, potential or apparent conflict of interest is known to us. I undertake to inform you of any change in these circumstances, including if an issue arises during the course of the meeting or work it (2023).

### References

- Burhan, A. (2009). *Second Language Teaching and Linguistics*. Palembang: Grafika Telindo Press.
- Gipps, C.V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London, UK: Routledge Falmer, 34-37
- Heaton, J.B. (1988). *Writing English Language Tests*. New York: Longman, Inc.
- Kubiszyn, T. & Borich, G. (1993). *Educational Testing and Measurement: Classroom Application and Practice*. New York: Harper Collins.
- Nguyen & Le. (2012). Communicative Language Testing: Do School Tests Measure Students' Communicative Competence, 856-864
- Oller, Jr., John W. (1979). *Language Tests at School*. London: Longman Group.