

## TEST DEVELOPMENT

(Disampaikan pada pelatihan guru bahasa Inggris se Kabupaten Sleman dalam rangka persiapan UAN yang diselenggarakan pada tanggal 29 Agustus 2008)

Jamilah, FBS, UNY  
Jamilah@uny.ac.id

### **Introduction**

A test, in simple terms is a method of measuring a person's ability, knowledge, or performance in a given domain. Let's look at the components of this definition. A test is first a method. It is an instrument- a set of techniques, procedures, or items,- that requires performance on the part of the test takers. To qualify as a test, the method must be explicit and structured: multiple choice questions with prescribed correct answers; a writing prompt with scoring rubric; an oral interview based on a question script and a checklist of expected responses to be filled in by the administrator.

Second, a test must measure. Some tests measure general ability, while others focus on a very specific competencies or objectives. Next, a test measures an individuals ability, knowledge, or performance. Testers need to know who the test takers are. What is their previous experience and background? Then, a test measures performance, but the result imply the test taker's ability or competence. Most language tests measure one's ability to perform language, that is to speak, write, read , and listen to a subset of language. On the other hand, it is not uncommon to find tests designed to tap into a test-taker's knowledge about language: defining a vocabulary item, reciting a grammatical rule, or identifying a rhetorical feature in written discourse. Finally, a test measures a given domain. In the case of a proficiency test, even though the actual performance on the test involves only a sample of skills, that domain is overall proficiency in a language- general competence in all skills of a language. Other tests may have very specific criteria. A test of pronunciation might be only a test of only a limited set of phonemic minimal pairs. A vocabulary test may focus on only the set of words covered in a particular lesson or unit. One of the biggest obstacles to overcome in constructing adequate tests is to measure the desired criterion and not include other factors inadvertently.

A well constructed test is an instrument that provides an accurate measure of the test-taker's ability within a particular domain. The definition may sound very simple, but in fact, constructing a good test is a complex task involving both science and art.

## **Types of tests**

There are many types of tests that teachers are likely to find during their life. They are aptitude test, proficiency tests, placement tests, diagnostic tests, and achievement tests.

A language aptitude test is designed to measure capacity or general capability to learn a foreign language and ultimate success in that undertaking. Language aptitude test is ostensibly designed to apply to the classroom learning of any language.

A proficiency test is a test designed to test global competence in a language. It is not limited to any one course, curriculum, or single skill in the language; rather, it tests overall ability. Proficiency tests have traditionally consisted of standardized multiple choice items on grammar, vocabulary, reading comprehension, and aural comprehension. Sometimes, a sample of writing is added, and more recent tests also include oral production performance. The example of standardized proficiency test is TOEFL.

Placemen test. Certain proficiency tests can act as placemen test. The purpose of placement test is to place a student into a particular level or section of a language curriculum or school. A placement test, usually but not always, includes a sampling of the materials to be covered in the various courses in the curriculum.

A diagnostic test is designed to diagnose specified aspects of a language. A test in pronunciation, for example, might diagnose the phonological features of English that are difficult for learners and should therefore become part of a curriculum. Usually such test offers a checklist of features for the administrator to use in pinpointing difficulties.

An achievement test is related directly to classroom lessons, units, or even a total curriculum. An achievement test should be limited to particular material addressed in a curriculum within a particular time frame and are offered after a course has focus in the objectives in question. Achievement tests can also serve the diagnostic role of indicating what a student needs to continue to work on in the future, but the primary role of an achievement test is to determine whether course objectives have been met, and appropriate knowledge and skills are acquired by the end of instruction.

## **Attributes of good test:**

Good tests must be valid, reliable, practical, and authentic. Validity is the most important principle of a test. Validity is the extent to which inferences made from test results are appropriate, meaningful, and useful in terms of the purpose of the test. A valid test of reading ability actually measures reading ability, not vision, nor previous knowledge in a subject, nor some other variables of questionable relevance. To see the validity of a test, there

is no final, absolute measure, but several different kinds of evidence may be evoked. In some cases, it may be appropriate to examine the extent to which a test calls for performance that matches that of the course or unit of study being tested. In other cases, we may be concerned of how well a test determines whether or not students have reached an established set of goals or levels of competence. Statistical correlation with other related but independent measures is another widely accepted form of evidence. Other concerns about a test's validity may focus on the consequences-of a test or even on the test-taker's perception of validity.

A reliable test is consistent and dependable. If we give the same test to the same student or matched students on different occasions, the test should yield similar results. The issue of reliability of a test may be addressed by considering a number of factors that may contribute to the unreliability of a test, the fluctuations in the students, in scoring, in test administration, and in the test itself.

A practical test is not excessively expensive, stays within appropriate time constraints, is relatively easy to administer, and has a scoring /evaluation procedure that is specific and time efficient. The value and quality of a test sometimes hinge on such nitty-gritty practical consideration.

The next attribute of good test is authenticity. Authenticity is the degree of correspondence of the characteristics of a given language test task to the features of a target language task. An authentic task is a task that is likely to be enacted in the real world. Many test items fail to simulate real world tasks. They may be contrived or artificial in their attempt to target a grammatical form or a lexical item. Test authenticity may be presented in the following: 1) The language in the test is as natural as possible, 2) Items are contextualized rather than isolated, 3) Topics are meaningful (relevant, interesting) for the learner, 4) Some thematic organization to items is provided, such as through a story line or episode, and 5) Tasks represent, or closely approximate, real world tasks.

### **Some Practical steps to test construction**

In preparing a test, there are 3 steps to follow: defining general purposes, drawing up test specifications, and devising test tasks or items.

a. Defining general purposes

Purposes of tests are usually in accordance with the type of test being constructed, whether it is an aptitude test, a proficiency test, placement test, diagnostic test, or an achievement test. It is better for us to know the natures of those tests.

b. Assessing clear, unambiguous objectives

In addition to know the purpose of the test you are creating, you need to know as specifically as possible what it is you want to test. You can do this by examining the objectives for the unit you are testing.

c. Drawing up test specification

Test specification for classroom use can be a simple and practical outline of your tests. For large-scale standardized tests that are intended to be widely distributed and therefore are broadly generalized, test specifications are much more formal and detailed. For class-use tests, your specifications will only comprise: a) a broad outline of the tests, b) what skills will you test, and c) what the items will look like. Specifying item format or types of items to be used is complex choices. These informal, classroom oriented specifications give you an indication of the topics (objectives you will cover, the implied elicitation and response formats for items, the number of items in each section, and the time to be allocated for each.

d. Devising test tasks

Devising test tasks or writing test items is the central part of test construction. Here you have to express and write what your students have to do to show what ability or competences you want to see. After writing your first draft of the task or test items, you need to ask yourself some important questions:

- 1) Are the directions to each section absolutely clear?
- 2) Is there an example item for each section?
- 3) Does each item measure a specified objective?
- 4) Is each item stated in clear, simple language?
- 5) Does each multiple choice item have appropriate distractors; that is, are the wrong items clearly wrong and yet sufficiently alluring that they aren't?
- 6) Is the difficulty of each item appropriate for your students?
- 7) Is the language of each item sufficiently authentic?
- 8) Do the sum of the items and the test as a whole adequately reflect the learning objectives?

After answering these question, when you find any answer NO you need to revise or improve your test.

Ideally, you need to try out all your test before actually administering the tests. But in our daily classroom teaching, the try out phase is almost impossible. Alternatively, you could enlist the aid of a colleague to look over your tests. And so you must do what you can to bring to your students an instrument that is, to the best of your ability, practical, and reliable..

## 2. Tryout of test items

If the test you construct is meant for standardized test, you must try out it. These are the components of the try out, what are being tried out and what needs to be done after the test has been tried out. The try outs cover the sample try out, item try out and analysis, student directions, administration direction, and time limit. Then we need to do technical analysis including estimating the test score means, estimating the standard deviation of test scores, and estimating reliability and item validity.

## Construction of the test items (objective tests: multiple choices)

### Suggestions for writing multiple choice items

1. The stem of multiple choice items should state a specific direct question. The stem should clearly formulate a problem.
2. Choose words that have precise meaning.
3. Avoid complex word arrangements
4. Adapt the range of difficulty of the item to the group and propose for which it is included.
5. Keep the options as short as possible
6. Use negative statements as sparingly as possible.
7. The distractors in a multiple choice item should be definitely incorrect but plausibly attractive to the weak students
8. Multiple choice items should provide at least four alternative answers.
9. Some of the most effective multiple choice items call for a best answer rather than an absolutely correct answer
10. Be sure no unintentional clues to the correct answers are given
11. Distractors should be related to the problems
12. Alternatives should be randomly ordered for each item.

## **WORKSHOP TASK: READING TEST ITEMS CONSTRUCTION**

To help you do the task in this training, that is, constructing Reading test items for use in the classroom, here are some materials that may be necessary for you to study:

## STANDAR KOMPETENSI LULUSAN

**LISTENING:** Memahami makna teks lisan berbentuk teks fungsional pendek, percakapan dan teks monolog sederhana berbentuk naratif (naratif, recount, news items) dan deskriptif (report, deskriptif, explanation) dalam konteks kehidupan sehari-hari.

**READING:** Memahami nuansa makna di dalam teks tertulis seperti teks fungsional pendek dan essay berbentuk naratif (naratif, recount, news item), deskriptif (report, deskriptif, explanation) dalam konteks kehidupan sehari-hari.

- : 1. menemukan gambaran umum, pikiran utama (overview questions)
2. Informasi tertentu (factual questions)
  3. informasi rinci
  4. informasi tersirat (inference questions)
  5. makna kata, makna frasa, makna kalimat
  6. Reference questions

## CONTOH-CONTOH PERTANYAAN

1. Menemukan gambaran umum (main topic, main idea, inti isi tulisan/judul):
  - What is the text about?
  - What is the main topic of the passage?
  - What is the topic of paragraph 1?
  - What is the main idea of the passage?
  - Which of the following is the best title of the passage?
  - What is the writer's purpose in writing the passage?
  - The first paragraph primarily describes...
2. Menemukan informasi tertentu (informasi yang tertera jelas dalam teks: tentang nama, tempat, tanggal, tahun, dsb).
  - What was Armstrong's nick name?
  - Where did Armstrong first learn to play the cornet?
  - In what city was .....?
  - In what period did.....? When .....

3. Menemukan informasi rinci (sama dengan informasi tertentu, biasanya dengan pertanyaan negative/yang tidak disebut dalam teks)
  - Which of the following requirements is NOT mentioned in the passage?
  - Which of the following can make your jewellery tarnish?
4. Menemukan informasi tersirat (informasi yang tidak tertera jelas dalam teks. Untuk dapat menjawab pertanyaan ini diperlukan kemampuan untuk menyimpulkan berdasarkan informasi yang ada)
  - Why does the author mention the experiments conducted several years ago?
  - The passage implies that, at birth, human babies....
  - It can be inferred from the passage that.....
  - The passage suggests that .....
5. Menemukan makna kata, frase, kalimat dalam teks (vocabulary in context)
  - The word ..... in line ..... is closest in meaning to
  - Which of the following is closest in meaning to the word .... In line .....?
  - The phrase.....in line ...is closest in meaning to which of the following?
6. Reference Questions
  - The word 'them' in line 4 refers to
  - To which of the following does the word 'these' in line 6 refer ?
  - The word 'these' in line 7 refers to which of the following?

====Good luck and Thank you====

#### References

Brown, H. Douglas, 2004. Language Assessment, Principles and Classroom Practices, Pearson Education, Longman