

Language Testing in the Classroom

EMMA B. MAGRACIA

Abstract

The three-part article focuses its discussion on the purposes for which tests are constructed with attention given to steps undertaken in test construction, the types of language tests, and samples of these tests with special attention on the language subskills (grammar and vocabulary) and language tests to measure the communication skills (listening, reading, speaking and writing).

Language teaching becomes meaningless without language testing. Testing the learners in their learning process is needed by the students and the teachers alike. The teachers will find out, through test results which serve as feedback, how far the students have imbibed the language lessons he/she has imparted to them. On the other hand, without tests, learning the language lessons becomes "boring" to the students. The students, as learners, become challenged and encouraged by test results. These test results serve as a gauge to the students to assess how far they have progressed in learning the language being taught to them. Hence, the importance played by language tests in the learning process taking place in the classroom is paramount in language teaching.

Purposes of Language Tests

Language tests, to be useful, must provide us with reliable and valid measurements for a variety of purposes, according to Henning (1987:1):

1. *Diagnosis and feedback.* This pinpoints strengths and weaknesses in the learned abilities of the student.



DR. EMMA B. MAGRACIA is a professor in Filipino and is presently designated as Coordinator of the Special General Education Night Program (SGENP) of the Institute. She had been designated likewise as Chair of the Filipino and Other Languages Department for several years, Assistant Dean, and Acting Dean of CASS.

2. *Screening and selection.* The result of this test is used to assist in the decision of who should be allowed to participate in a particular program of instruction.
3. *Placement.* Tests are used to identify a particular performance level of the student and to place him/her at an appropriate level of instruction.
4. *Program evaluation.* This provides information about the effectiveness of programs of instruction. The focus of the evaluation is not the individual student but the actual program of instruction.
5. *Providing research criteria.* This provides a standard of judgment in a variety of other research contexts. Comparisons of methods and techniques of instruction, textbooks, or audiovisual aids usually entail reference to test scores.
6. *Assessment of attitudes and socio-psychological differences.* Attitudes toward the target language, its people, and their culture have been identified as important affective correlates of good language learning (Naiman et al., 1978). Apart from attitudes, other variables such as cognitive style of the learner (Witkin et al., 1977), socioeconomic status, linguistic situational context (Henning, 1979), and ego-permeability of the learner (Henning, 1979) have been found to relate to levels of language achievement and/or strategies of language use.

The above enumerated tests are of wider application and are prepared by test experts and psychometricians. The results of these tests are bases for decision making for the effective implementation of established policies and institutional practices.

Types of Language Tests

Just as there are many purposes for which language tests are developed, so there are many types of language tests. Some types of tests serve a variety of purposes while others are more restricted in their applicability. There are, however, many important broad categories of tests that do permit more efficient description and explanation. Many of these categories stand in opposition to one another while some are merely mental constructs to facilitate understanding.

1. *Objective vs. subjective tests.* Usually these types of tests are distinguished on the basis of the manner in which they are scored. An objective test is one that may be scored by comparing examinee responses with an established set of acceptable responses or scoring key. No particular knowledge or training in the examined content area is required on the part

of the scorer. A common example would be a multiple-choice recognition test. Conversely, a subjective test is said to require scoring by opinionated judgment, based on insight and expertise, on the part of the scorer. An example might be the scoring of free, written compositions where no operational definitions of creativity are provided and where there is only one rater. Many tests, such as cloze tests permitting all grammatically acceptable responses to systematic deletions from a context, lie somewhere between the extremes of objectivity and subjectivity (Oller, 1979).

2. *Direct vs. indirect tests.* It has been said that certain tests, such as ratings of language use in real and uncontrived language situations, are testing language performance directly, whereas other tests, such as multiple-choice recognition tests, are indirectly tapping true language performance and therefore are less valid for measuring language proficiency. Many language tests are viewed as lying somewhere on a continuum from natural-situational to unnatural-contrived. Thus an *interview* may be thought of as a more direct than a *cloze* test for measuring overall language efficiency. A contextualized vocabulary test is also thought to be more natural and direct than synonym-matching test.
3. *Discrete-point vs. integrative tests.* In 1961, John B. Carroll made this distinction between these types of tests. Discrete-point tests are designed to measure knowledge or performance in very restricted areas of the target language. An example is a test of ability to use correctly the perfect tenses of English verbs or to supply correct prepositions in a cloze passage. On the other hand, integrative tests tap a greater variety of language abilities concurrently and therefore may have less diagnostic and remedial-guidance value and greater value in measuring overall language proficiency. Examples are random cloze, dictation, oral interviews, and oral imitation tasks.
4. *Aptitude, achievement and proficiency tests.* *Aptitude tests* are most often used to measure the suitability of a candidate for a specific program of instruction or a particular kind of employment. A language aptitude test may be used to predict the likelihood of success of a candidate for instruction in a foreign language. Frequently, vocabulary tests are effective aptitude measures. *Achievement tests* are used to measure the extent of learning in a prescribed content domain. These tests which are also known as *attainment tests*, are concerned with assessing what has been learned of a known syllabus within a school or within a total educational system. These are administered to find out just how much has been learned of what has been taught (Davies, 1977). Such tests normally come after a program of instruction and that the components or items of the tests are drawn from the content of instruction directly (Mehrens and Lehmann, 1975). If the

purpose of achievement testing is to isolate learning deficiencies in the learner with the intention of remediation, such tests may also be called *diagnostic tests*. *Proficiency tests* are most often global measures of ability in a language or other content area. These measures are often used for placement or selection. It may be noted that the primary differences among these three kinds of tests are in the purposes they serve and the manner in which their content is chosen.

5. *Criterion- or domain-referenced vs. norm-referenced or standardized tests*. There is no essential difference between *criterion-referenced tests* and *domain-referenced tests* (Hambleton, et al., 1978). *Criterion-referenced tests* are devised before instruction itself is designed. The test must match teaching objectives perfectly so that any tendency of the teacher to "teach to the test" would be permissible. A criterion or cut-off score is set in advance (usually 80 to 90 percent of the total possible score) and those who do not meet the criterion are required to repeat the course. Students are evaluated in terms of their relative degree of mastery of course content, rather than with respect to their relative ranking in the class. Thus, all students who master the course content might receive an 'A' irrespective of how many students achieve this grade. A *norm-referenced or standardized test* has been previously administered to a large sample of people from the target population (e.g., 1,000 or more). Test results are interpreted with reference to the performance of a given group, or norm. This is designed and developed to maximize distinctions among individual test takers. This means that the items or parts of such tests will be selected according to how well they discriminate individuals who do well on the test as a whole from those who do poorly (Bachman, 1990). Acceptable standards of achievement can only be determined after the test has been developed and administered. Such standards are found by reference to the mean or average score of other students from the same population.
6. *Speed tests vs. power tests*. A purely *speed test* is one in which the items are so easy that every person taking the test might be expected to get every item correct, given enough time. But sufficient time is not provided, so examinees are compared on their speed of performance rather than on knowledge alone. Conversely, *power tests* by definition are tests that allow sufficient time for every person to finish, but that contain such difficult items that few if any examinees are expected to get every item correct. Most tests fall somewhere between the two extremes since knowledge rather than speed is the primary focus, but time limits are enforced since weaker students may take unreasonable periods of time to finish.

7. *Productive skills tests vs. receptive skills tests.* These tests generally assess the four macro skills of an examinee. The tests are classified to measure the speaking and writing skills that require active or creative answers like the speaking examinations (productive) and the reading and listening skills which tend to rely on recognition like the multiple-choice reading tests (receptive).
8. *Other test categories.* Some other test categories that might be mentioned are *examinations vs. quizzes, questionnaires, and rating schedules.* Then mention could be made of *cloze tests, dictation tests (especially dicto-comp tests), maze, multiple-choice tests, true/false tests, etc.*

Phases in Test Construction and Specifications

Carroll (1980) has identified three phases in the construction of tests which are:

- | | |
|-----------------------------|-------------------------------------|
| Phase 1: <i>Design</i> | 1. Description of participant(s); |
| | 2. Analysis of communicative needs; |
| | 3. Specification of test content; |
| Phase 2: <i>Development</i> | 4. Realization of tests; |
| | 5. Trial application; |
| | 6. Validation and test analysis; |
| Phase 3: <i>Operation</i> | 7. Full scale application; |
| | 8. Operational use; |
| | 9. Revision of test system. |

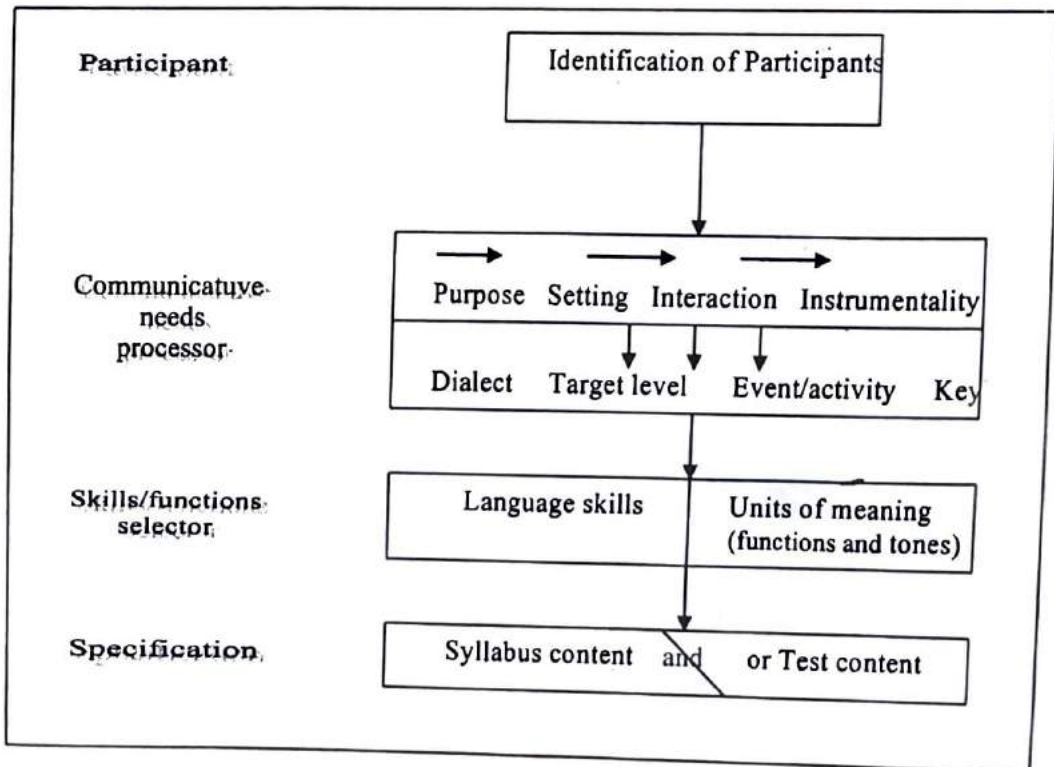
As may be noted in the above identification by Carroll, the construction of tests undergo three important phases, namely the designing phase, the development phase, and the operation phase. Although Carroll seems to present these phases in a linear form, in actuality test construction is cyclical in nature. Any test, especially those with wider application, has to be revised and improved to serve its purpose fully.

The test design has to conform with the participants' needs particularly to measure their skills in communication. Rapid social and political changes demand the need for a new test or assessment procedure. These new situations include the growth of international education, increased labor flows between countries, the educational impact of immigration and refugee programs, school curriculum reforms due to varied impacts of globalization, and many others. Given these multifarious conditions, the design of the test to be constructed should consider the participants

and their communicative needs so test content specification can cover the scope that it has to measure. Establishing the test content can involve careful sampling from the domain of the test, the set of tasks or kind of behavior in the criterion setting of the test construct. The test domain may define a set of practical, real-world tasks or in terms of knowledge of the grammatical system, or of vocabulary, or of features of pronunciation, or ability to perform aspects of the skill areas of speaking, listening, reading or writing.

The next phase has to do with the actual realization of tests, the trial application these tests have to undergo so that the validation and analysis of the test items can materialize. This stage can be compared to the making of a car where after it has been designed and manufactured, has to undergo a road test to find out its efficiency and overall applicability. An ordinary classroom test may not need trial testing or trying out with a trial population, a set of people who resemble in all relevant characteristics the target test population. However, for some kinds of tests with a wider application, this stage is an important phase that has to be accomplished to gather feedback which will be relevant to the possible revision and improvement of the test.

The last phase will need the full scale application of the test, its operational use in the fields where it has applicability and the expected revision of the test after it has been applied fully. Data from actual test performance needs to be systematically gathered and analyzed, and to investigate the validity and usefulness of the test under operational conditions. Periodically, the results of these data will lead to



substantial test revision and the cycle of the phases of test construction will recommence.

A simpler presentation of the parameters and procedures for test design as presented above is further explained by Carroll (1980:24). The details for each particular area is specifically given.

Information Banks	Procedural Guide
1 Participant identification	Broadly describe typical participant.
2 Purpose for language use	Describe main uses of English and classify under ESP headings: academic, occupational or social survival.
3 Events/activities	Choose the major events to be met with, and select several activities for each.
4 Instrumentality	Select media – listening, speaking, reading or writing, or multiple-mode combinations. Channels: face-to-face, tape, print, film, etc.
5 Socio-Cultural	Specify social relationships, dialect and socio-cultural factors.
6 Performance levels	Using nine-point scale, give target levels of performance for each medium and multiple-mode.
7 Topic areas	Identify semantic areas for each specified event.
8 Language skills	Choose skills necessary for carrying out the different activities at given target levels.
9 Language function/tones units	Indicate functions needed, and appropriate attitudinal tones, for those activities involving sizeable person-to-person interactions.

10 Test format

Choose types of item for each activity – closed-ended, open-ended or restricted response (note RACE and authenticity)

The schematic presentations above are for tests of wider application. For classroom tests, a simpler test design specification is normally made. A crude model of a test design specification which is commonly used by teachers is the following:

Levels	Skills	Production		Reception	
		Writing	Speaking	Reading	Listening
Phonology					
Vocabulary					
Grammar					
Context					

The figure above is a simpler production of test specification which the teacher uses. Here the items are defined and clearly indicated so a balance of the skills and aspects of language assessment can be covered. It may be noted that language levels are specified and the language skills determined so the levels and skills to be tested will be properly covered in the scope of the test construction.

Sample Language Tests

Following are sample language tests. However, only a very limited number is presented for the purpose of simplicity and exemplification. The samples given are divided into testing language subskills (vocabulary and grammar tests) and testing communication skills (reading, writing, listening, and speaking). Most of the sample tests are taken from illustrations given by Madsen (1983).

Vocabulary tests. These tests are used to measure the comprehension and production of words used in speaking or writing. The most simple is the *limited response* for beginners. This test item requires either a simple physical action like pointing at something or a very simple verbal answer such as yes or no. The *multiple-choice completion* is a test in which a sentence with a missing word is presented and the student chooses one of four vocabulary items given to complete the sentence. The *multiple-choice paraphrase* is a test in which a sentence with one word

underlined is given. The examinee chooses which of four words is the closest in meaning to the underlined item. The fourth kind is the *simple completion* (words) where the student writes the missing part of words that appear in the sentence.

1. Limited response

Ex. Please go to the window. (The examinee goes to the window.) Or What color is the book? (Green or whatever color the book is)

2. Multiple-choice completion – the examinee depends on context clues and sentence meaning to get the correct answer

Ex. She quickly _____ her lunch.

A. drank *B. ate C. drove D. slept

Or a mini dialog as

“I want to paint , too.”

“All right. Use that _____ over there.”

*A. brush B. pencil C. broom D. spoon

Or you can use *multiple-choice cloze*. *Cloze tests* are made from stories or essays by deleting words at regular intervals. They provide more context – often more than one paragraph. Multiple-choice cloze can test vocabulary when only content words are deleted, in which case deletion does not follow regular intervals.

Ex. After the capture of Troy, Ulysses set out for his

(A. neighborhood B. continent *C. homeland D. street) many miles

away. But so many strange (A. sights *B. things C. places D. people)

happened to him on his journey that ten (*A. years B. times C. roads

D. cities) passed before he reached Ithaca!

3. Multiple-choice paraphrase – choosing the right word depends more on knowing the key vocabulary item than on finding meaning in the sentence.

Ex. He was **irate** when he heard about the new plans.

A. interested B. surprised *C. angry D. sad

Or My sister is a **pilot**. She can _____.

A. help sick people B. make clothes *C. fly an airplane D. teach students at school

4. Simple completion (words) – these require students to fill in missing parts of words that appear in sentences which are usually prefixes and suffixes.

Ex. **un-** in **untie** or the **-ful** in **thankful**.

Ex. (beauty) She has a _____ new dress. *beautiful

Grammar tests. Grammar tests are designed to measure student proficiency in matters ranging from inflections (bottle—bottles, bake—baked) to syntax. Syntax

involves the relationship of words in a sentence, including matters such as word order, use of the negative, question forms, and connectives.

Ex. She is _____ her breakfast.

*A. eating B. ate C. eats D. eaten

Or He _____ (sleep) well when he was a child.

Or Complete the following sentences with "do" or "make."

1. He _____ a lot of money last year. *made

2. I always _____ my best. *do

Or He's the _____ (tall) person in the class. *tallest

Communication skills tests measure actual communication in the real world. Generally, these tests measure integrative skills and combine elements like vocabulary and grammar as well as matters beyond the sentence level, such as writer's purpose or even social appropriateness.

Reading tests. Tests of reading come in a wide variety of forms and evaluate a broad spectrum of reading activities. These range from pre-reading concerns to reading comprehension, reading speed, and skimming techniques and also advanced and more specialized applications that include translation, reading aloud and reading literature. The heart of reading evaluation in most schools is reading comprehension, and that will be the emphasis in this paper.

One of the simplest forms of sentence comprehension is the true-false sentence for beginners. The student looks for truths, untruths or impossibilities.

Ex. The sun sets in the east. T *F

She is my brother. T *F

Use a variety of types of questions on your reading test. One very important type is the *paraphrase*. Look at the following example.

Karate is a science of unarmed self-defense and counterattack. It is a sort of "weapon in an empty hand." In many U.S. cities thousands of young people are developing their minds as well as their bodies by learning karate.

In this passage we learn that karate _____

*A. is being taught to many young Americans.

B. and training for the mind are both being taught.

C. can remove a weapon from someone's hand.

D. is used to start a fight.

The second type of question, the *synthesis* item, requires integration of ideas from more than one sentence – sometimes from the entire selection. Here is the full version of the karate passage:

Karate is a science of unarmed self-defense and counterattack. It is a sort of "weapon in an empty hand." In many U.S. cities thousands of young people are developing their minds as well as their bodies by learning karate. "I've been taking karate lessons for five years now," says sixteen-year-old Bobby Hamilton of Columbus, Ohio, "and it's great! I find myself doing things that I thought I could never do." Paula Jones has just begun taking karate lessons at her high school in Philadelphia. She feels that she has more self-confidence because of the lessons. "I am more aware of myself," she says. "I already have learned so much about self-control. I know everything in life is not going to be easy. Karate helps prepare me for the times when I'll have to meet my problems face to face."

A good title for this selection would be _____.

- A. Americans Import a Japanese Sport
- B. Karate – Weaponless Protection for People of All Ages
- C. School Children Enjoy a New Kind of Physical Education Class
- *D. Self-Perfection Through Self-Protection

A third kind of question is the *inference item*. It requires students to see implications in what they read. A sample follows.

[Two men, Gerard and Denys, were traveling in a forest. They had just been forced to kill a large baby bear in self-defense.] Then Gerard heard a sound behind them. It was a strange sound, too, like something heavy, but not hard, rushing over dry leaves. It was a large bear, as big as a horse, running after them a short distance away. As soon as he saw it, Gerard cried out in fear, "The baby bear's mother."

The mother bear was probably running because it _____.

- A. was afraid of Gerard and Denys and wanted to escape.
- *B. wanted to hurt those who had killed its baby.
- C. was chasing a horse, a short distance away.
- D. Enjoyed running, like horses and other animals do.

Multiple-choice cloze can likewise be used in this test type using a passage. In addition, there are alternate forms of passage-comprehension items that require active rather than passive responses. One example is the sentence-level item which is extremely easy to prepare, and questions requiring only simple responses minimize the mixing of writing skills and reading skills. An example would be:

Gerard cried out in fear because _____.

Another alternative is the editing test. The *editing* test is a kind of cloze in reverse. Instead of taking words out, we put words in. Usually these are common words used in the textbook. They must not be put in at fixed intervals.

The boy was called Abe. He lived read in a cabin on a small farm. Abe lived with his mother and father stores and sister. His clothes were homemade. He didn't even own language a pair of good shoes. But spend he never went hungry or cold. Abe had to work hard on need the farm, but he was happy because wild there was a lot of love in his family.

*read, stores, language, spend, need, wild

Another alternative is the test that checks reading speed. A special problem in measuring reading speed is being sure that the material has been understood. After having students do a speed-reading check, some teachers give a simple true-false test. Students must have something like 80 percent comprehension for full credit on the speed test.

Writing tests. There are many kinds of writing tests. The reason for this is simple: a wide variety of writing tests is needed to test the many kinds of writing tasks that we engage in. There are usually distinct stages of instruction in writing, such as pre-writing, guided writing, and free writing. Each stage tends to require different types of evaluation. Our discussion will focus on evaluating writing in its various stages, from pre-writing to free writing with emphasis on guided writing including techniques such as dictation.

Some techniques used in evaluating pre-writing stage of writing instruction are sentence combining, sentence expansion, sentence reduction and copying which may be timed. One copying test uses the maze type which can make students more aware of extended discourse and also of mechanics especially punctuation and spelling.

Ex.

(Yesterday,)	(friends)	(her)
(Last week,) one of my	(teachers) asked me to help	(him) plan
(Recently,)	(brothers)	(_)
(an activity)	(boys)	(class.)
(a party) for the	(girls) in our	(dorm.) etc....
(a social)	(students)	(apartment.)

The objective in guided-writing tests is to check student ability to handle controlled or directed writing tasks. The techniques employed are text manipulation(to make certain kinds of changes in a story), expand the outline of an article, and dictation.

An old but still acceptable way of checking spelling is to dictate problem words orally. Another way to check spelling is to use a multiple-choice arrangement.

Ex. *Directions:* In each of the following groups of four words, one word is spelled incorrectly; the other three are spelled correctly. Find which word is spelled wrong. Then circle the letter of that word.

- A. believe B. all right C. because *D. marriage

Unpunctuated sentences such as the following can be used to check both punctuation and capitalization.

Ex. on december 25 1980 doctor adams died in an automobile accident nevertheless his kind deeds will live on for a long time

*On December 25, 1980, Dr. Adams died in an automobile accident; nevertheless, his kind deeds will live on for a long time.

Another guided writing test may use a paragraph outline, use of topic sentence, or guide questions.

Another guided-writing test is dictation. Most teachers know about this technique but few handle it properly. Actually this is one of the easiest tests to use, and it gives very good information on the student's language ability. The most common tests are the simple dictation, dicto-comp, and dicto-gloss. In the dictation test, the passage or text is read by the teacher three times. In the first reading, the students are to listen only and are not allowed to write anything. In the second reading, they are to write the text as dictated, and in the third dictation the students may make corrections in their work.

In the dicto-comp, the passage is read four times. The students are not allowed to write sentences but may write down some important words. After the fourth reading, they will write down the entire composition trying to follow the original text. They may use substitute words but care must be made to retain the original idea of the composition.

Dicto-gloss is almost the same as dicto-comp. The only difference is that in dicto-gloss, the students are not allowed to take down cue words from the text being read by the teacher. They have to rely on what they have heard and from their memory to compose the text.

Few teachers have students write without giving them a specific topic. We need to be certain that each student is performing the same kind of task. Otherwise, we cannot make a fair comparison of their writing. The aim in a writing test is generally to evaluate the effectiveness of the total composition including sentence-level accuracy, larger rhetorical matters such as unity, coherence, and organization, as well as effectiveness in conveying ideas to the intended audience, including socially appropriate language and appropriate selection of supporting details. Techniques that can be used are a picture sequence that tells a story which the

students can make a narrative on, use of a chart or table or diagram for the students to interpret or use of a very specific situation to serve as guideline for your student writing.

Listening tests. There are two broad categories of tests that incorporate the listening skills. One group simply uses listening as a tool to evaluate something else such as responding to simple commands or a means of evaluating low-level proficiency in grammar and pronunciation. Another group is used to evaluate more advanced integrative skills, those that evaluate proficiency in the listening skill itself. These include the recognition of words and structures and pronunciation features, the difference between subskill tests using listening as a tool and the integrative listening comprehension.

One type of test used in testing listening skill is the multiple-choice response. Three things must be kept in mind in preparing such test: focus on meaning, keep the options simple, and learn to adjust the difficulty of the items.

Ex. (When Jack leaves, they'll hire you, won't they?) (This is heard from the teacher)

A. Yes, you will. B. Yes, he's leaving. *C. Yes, they will. (This part is read)

Ex. (Bert isn't any heavier than George.) (This is heard from the teacher.)

A. Bert is heavier. B. George is heavier. *C. They weigh about the same.

Another method is reading a passage or a lecture and taking notes on it, and the students would answer printed questions as the following:

Ex. Where is Kochen? A. In Europe B. In Central America *C. In the Far East

What is the normal range of temperature in Kochen?

*A. Close to 40 degrees B. About 50 degrees C. Around 60 to 65 degrees

Other techniques that may be used are the use of dialogs, paraphrase options, taped radio or TV commercials, or use of situations involving an inquiry or a routine transaction.

Speaking tests. The testing of speaking is widely regarded as the most challenging. The nature of speaking skill itself is not usually well defined. Grammar, vocabulary, and pronunciation are often named as ingredients but fluency and appropriateness of expression are usually regarded as equally important. Still other factors identified in oral communication include listening comprehension, correct tone (sadness or fear, for instance), reasoning ability, and such things as initiative in asking for clarification. Some techniques that are quite effective are the use of picture-cue items, reading-aloud passage, paraphrase, explanation, guided role play, and oral interview.

Conclusion

Apparently, language teaching becomes more meaningful, challenging, and gratifying when feedbacks are regularly inputted through test results. Both students and teachers become intensely aware of the rate they are receiving and giving in terms of language lessons and practical communication activities. Thus, there is no substitute for the role played by language tests in a classroom setting.

References

- Alderson, J. Charles. 1991. "Bands and Scores." J. Charles Alderson and B. North, eds. **Language Testing in the 1990s**. London: Modern English Publications and the British Council.
- Bachman, Lyle F. 1990. **Fundamental Considerations in Language Testing**. New York: Oxford University Press.
- _____ and Adrian S. Palmer. 1996. **Language Testing in Practice: Designing and Developing Useful Language Tests**. New York: Oxford University Press.
- Carroll, Brendan J. 1980. **Testing Communicative Performance**. Oxford: Pergamon Press.
- Davies, Alan. 1977. "The Construction of Language Tests." **The Edinburgh Course in Applied Linguistics Volume 4: Testing and Experimental Methods**. Oxford: Oxford University Press.
- Hanson, F. Allan. 1993. **Testing Testing: Social Consequences of the Examined Life**. California: University of California Press.
- Henning, Grant. 1987. **A Guide to Language Testing: Development, Evaluation, Research**. Rowley, Mass.: Newbury House.
- Larsen-Freeman, Diane and Michael H. Long. 1991. **An Introduction to Second Language Acquisition Research**. London: Longman.

- Madsen, Harold S. 1983. **Techniques in Testing**. New York: Oxford University Press.
- McNamara, Tim. 2000. **Language Testing**. Oxford: Oxford University Press.
- _____. 1996. **Measuring Second Language Performance**. London: Longman.
- Magracia, Emma B. "Ang Pagtuturo ng Wika at ang Pagsusulit Pangwika." *Kawika*. Bol. I, Blg. 1 & 2. (Hunyo-Nobyembre 1982), 33-39.
- Naiman, N., et al. 1978. "The Good Language Learner." *Research in Education Series*. No. 7. Toronto: The Ontario Institute for Studies in Education.
- Oller, John W. 1979. **Language Tests at School**. New York: Longman.
- _____, ed. 1983. **Issues in Language Testing Research**. New York: Longman.
- Omaggio, Alice C. 1986. **Teaching Language in Context**. Boston, Mass.: Heinle & Heinle Publishers, Inc.
- Santos, Angelina L. "Ang mga Pagsusulit na Pangwika." *Kawika*. Bol. II, Blg. 1 (Hunyo 1983), 99-121.
- Shohamy, Elana. 1985. **A Practical Handbook in Language Testing for the Second Language Teacher**. Tel Aviv: Tel Aviv University.
- Spolsky, Bernard. 1995. **Measured Words**. Oxford: Oxford University Press.
- Weir, Cyril J. 1990. **Communicative Language Testing**. New York: Prentice Hall.
- _____. 1993. **Understanding and Developing Language Tests**. New York: Prentice Hall.
- Witkin, H. A., et al. 1971. **A Manual for the Embedded Figures Test**. Palo Alto, Cal.: Consulting Psychologists Press.