

Introduction to Issues in Language Assessment and Terminology

In today's language classrooms, the term *assessment* usually evokes images of an end-of-course paper-and-pencil test designed to tell both teachers and students how much material the student doesn't know or hasn't yet mastered. However, assessment is much more than tests. Assessment includes a broad range of activities and tasks that teachers use to evaluate student progress and growth on a daily basis.

Consider a day in the life of Ms. Wright, a typical experienced ESL teacher in a large urban secondary school in Florida. In addition to her many administrative responsibilities, she engages in a wide range of assessment-related tasks on a daily basis. It is now May, two weeks before the end of the school year. Today, Ms. Wright did the following in her classroom:

- graded and analyzed yesterday's quiz on the irregular past tense
- decided on topics for tomorrow's review session
- administered a placement test to a new student to gauge the student's writing ability
- met with the principal to discuss the upcoming statewide exam
- checked her continuous assessment records to choose students to observe for speaking today
- improvised a review when it was clear that students were confused about yesterday's vocabulary lesson
- made arrangements to offer remediation to students who did poorly on last week's reading practice exam
- after reviewing the final exam that came with the textbook, decided to revise questions to suit class focus and coverage
- graded students' first drafts of a travel webquest using checklists distributed to students at the start of the project

Each of these tasks was based on a decision Ms. Wright made about her students or her class as a whole. Teachers assess their students in a number of ways and for a variety of purposes because they need to make decisions about their classrooms and their teaching. Some of these decisions are made on the

spot, such as the improvised review. Others, like preparing the final exam, entail long-term planning.

Placing students in the right level of classroom instruction is an essential purpose of assessment. Normally, new students are given placement exams at the beginning of the school year, but some new students arrive throughout the year. By assigning a new student a writing task to gauge her writing ability, Ms. Wright tried to ensure that the student would benefit from instruction at the appropriate level for the remaining weeks of the school year.

Some of the decisions Ms. Wright made today had to do with **diagnosing student problems**. One of a teacher's main aims is to identify students' strengths and weaknesses with a view to carrying out revision or remedial activities. By making arrangements to offer remediation to students who did poorly on last week's reading exam, she was engaging in a form of **diagnostic assessment**.

Much of what teachers do today in language classrooms is to **find out about the language proficiency of their students**. In preparing her students to take the Florida Comprehensive Assessment Test (FCAT), Ms. Wright was determining whether her students have sufficient language proficiency to complete the exam effectively and meet national benchmarks.

Other activities were carried out with the aim of **evaluating academic performance**. In fact, a lot of teacher time is spent gathering information that will help teachers make decisions about their students' achievement regarding course goals and mastery of course content. Ms. Wright uses multiple measures such as quizzes, tests, projects, and continuous assessment to monitor her students' academic performance. To assign speaking grades to her students, she had to select four or five students per day for her continuous assessment records. These daily speaking scores will later be averaged together with her students' formal oral interview results for their final speaking grades.

Many of her classroom assessment activities concerned **instructional decision-making**. In deciding which material to present next or what to revise, Ms. Wright was making decisions about her language classroom. When she prepares her lesson plans, she consults the syllabus and the course objectives, but she also makes adjustments to suit the immediate needs of her students.

Some of the assessment activities that teachers participate in are for **accountability purposes**. Teachers must provide educational authorities with evidence that their intended learning outcomes have been achieved. Ms. Wright understands that her assessment decisions impact her students, their families, her school administration, and the community in which she works.

Evaluation, Assessment, and Testing

To help teachers make effective use of evaluation, assessment, and testing procedures in the foreign/second (F/SL) language classroom, it is necessary to clarify what these concepts are and explain how they differ from one another.

The term *evaluation* is all-inclusive and is the widest basis for collecting information in education. According to Brindley (1989), evaluation is "conceptualized as broader in scope, and concerned with the overall program" (p. 3). Evaluation involves looking at all factors that influence the learning process, i.e., syllabus objectives, course design, and materials (Harris & McCann, 1994). Evaluation goes beyond student achievement and language assessment to consider all aspects of teaching and learning and to look at how educational decisions can be informed by the results of alternative forms of assessment (Genessee, 2001).

Assessment is part of evaluation because it is concerned with the student and with what the student does (Brindley, 1989). *Assessment* refers to a variety of ways of collecting information on a learner's language ability or achievement. Although *testing* and *assessment* are often used interchangeably, *assessment* is an umbrella term for all types of measures used to evaluate student progress. *Tests* are a subcategory of assessment. A *test* is a formal, systematic (usually paper-and-pencil) procedure used to gather information about students' behavior.

In summary, *evaluation* includes the whole course or program, and information is collected from many sources, including the learner. While *assessment* is related to the learner and his or her achievements, *testing* is part of assessment, and it measures learner achievement. //

Categorizing Assessment Tasks

Different types of tests are administered for different purposes and used at different stages of the course to gather information about students. You as a language teacher have the responsibility of deciding on the best option for your particular group of students in your particular teaching context. It is useful to categorize assessments by type, purpose, or place within the teaching/learning process or timing.

Types of Tests

The most common use of language tests is to identify strengths and weaknesses in students' abilities. For example, through testing we might discover that a student has excellent oral language abilities but a relatively low level of reading comprehension. Information gleaned from tests also assists us in deciding who should be allowed to participate in a particular course or program area. Another common use of tests is to provide information about the effectiveness of programs of instruction.

Placement Tests

Placement tests assess students' level of language ability so they can be placed in an appropriate course or class. This type of test indicates the level at which a student will learn most effectively. The primary aim is to create groups of learners that are homogeneous in level. In designing a placement test, the test developer may base the test content either on a theory of general language proficiency or on learning objectives of the curriculum. Institutions may choose to use a well-established proficiency test such as the TOEFL®, IELTS™, or MELAB exam and link it to curricular benchmarks. Alternatively, some placement tests are based on aspects of the syllabus taught at the institution concerned (Alderson, Clapham, & Wall, 1995).

At some institutions, students are placed according to their overall rank in the test results combined from all skills. At other schools and colleges, students are placed according to their level in each skill area. Additionally, placement test scores are used to determine if a student needs further instruction in the language or could matriculate directly into an academic program without taking preparatory language courses.

Aptitude Tests

An *aptitude test* measures capacity or general ability to learn a foreign or second language. Although not commonly used these days, two examples deserve mention: the Modern Language Aptitude Test (MLAT) developed by Carroll and Sapon in 1958 and the Pimsleur Language Aptitude Battery (PLAB) developed by Pimsleur in 1966 (Brown, H.D., 2004). These are used primarily in deciding to sponsor a person for special training based on language aptitude.

Diagnostic Tests

Diagnostic tests identify language areas in which a student needs further help. Harris and McCann (1994) point out that where "other types of tests are based

on success, diagnostic tests are based on failure" (p. 29). The information gained from diagnostic tests is crucial for further course activities and providing students with remediation. Because diagnostic tests are difficult to write, placement tests often serve a dual function of both placement and diagnosis (Harris & McCann, 1994; Davies et al., 1999).

Progress Tests

Progress tests measure the progress that students are making toward defined course or program goals. They are administered at various stages throughout a language course to determine what students have learned, usually after certain segments of instruction have been completed. *Progress tests* are generally teacher produced and narrower in focus than achievement tests because they cover less material and assess fewer objectives.

Achievement Tests

Achievement tests are similar to progress tests in that they determine what a student has learned with regard to stated course outcomes. They are usually administered at mid- and end-point of the semester or academic year. The content of achievement tests is generally based on the specific course content or on the course objectives. *Achievement tests* are often cumulative, covering material drawn from an entire course or semester.

Proficiency Tests

Proficiency tests, on the other hand, are not based on a particular curriculum or language program. They assess the overall language ability of students at varying levels. They may also tell us how capable a person is in a particular language skill area (e.g., reading). In other words, proficiency tests describe what students are capable of doing in a language.

Proficiency tests are typically developed by external bodies such as examination boards like Educational Testing Services (ETS), the College Board, or Cambridge ESOL. Some proficiency tests have been standardized for international use, such as the TOEFL®, which measures the English language proficiency of foreign college students who wish to study in North American universities or the IELTS™, which is intended for those who wish to study in the United Kingdom or Australia (Davies et al., 1999). Increasingly, North American universities are accepting IELTS™ as a measure of English language proficiency.

Additional Ways of Labeling Tests

Objective versus Subjective Tests

Sometimes tests are distinguished by the manner in which they are scored. An *objective test* is scored by comparing a student's responses with an established set of acceptable/correct responses on an answer key. With objectively scored tests, the scorer does not require particular knowledge or training in the examined area. In contrast, a *subjective test*, such as writing an essay, requires scoring by opinion or personal judgment so the human element is very important.

Testing formats associated with objective tests are multiple choice questions (MCQs), True/False/Not Given (T/F/Ns), and matching. Objectively scored tests are ideal for computer scanning. Examples of subjectively scored tests are essay tests, interviews, or comprehension questions. Even experienced scorers or markers need moderated training sessions to ensure inter-rater reliability.

Criterion-Referenced versus Norm-Referenced or Standardized Tests

Criterion-referenced tests (CRTs) are usually developed to measure mastery of well-defined instructional objectives specific to a particular course or program. Their purpose is to measure how much learning has occurred. Student performance is compared only to the amount or percentage of material learned (Brown, J.D., 2005).

True CRTs are devised before instruction is designed so that the test will match the teaching objectives. This lessens the possibility that teachers will "teach to the test." The criterion or cut-off score is set in advance. Student achievement is measured with respect to the degree of learning or mastery of the pre-specified content. A primary concern of a CRT is that it be sensitive to different ability levels.

Norm-referenced tests (NRT) or standardized tests differ from criterion-referenced tests in a number of ways. NRTs are designed to measure global language abilities. Students' scores are interpreted relative to all other students who take the exam. The purpose of an NRT is to spread students out along a continuum of scores so that those with low abilities in a certain skill are at one end of the normal distribution and those with high scores are at the other end, with the majority of the students falling between the extremes (Brown, J.D., 2005, p. 2).

By definition, an NRT must have been previously administered to a large sample of people from the target population. Acceptable standards of achievement are determined after the test has been developed and administered. Test results are interpreted with reference to the performance of a given group or

norm. The *norm* is typically a large group of students who are similar to the individuals for whom the test is designed.

Summative versus Formative

Tests or tasks administered at the end of the course to determine if students have achieved the objectives set out in the curriculum are called *summative assessments*. They are often used to decide which students move on to a higher level (Harris & McCann, 1994). *Formative assessments*, however, are carried out with the aim of using the results to improve instruction, so they are given during a course and feedback is provided to students.

High-Stakes versus Low-Stakes Tests

High-stakes tests are those in which the results are likely to have a major impact on the lives of large numbers of individuals or on large programs. For example, the TOEFL® is high stakes in that admission to a university program is often contingent on receiving a sufficient language proficiency score.

Low-stakes tests are those in which the results have a relatively minor impact on the lives of the individual or on small programs. In-class progress tests or short quizzes are examples of low-stakes tests.

Traditional versus Alternative Assessment

One useful way of understanding alternative assessment is to contrast it with traditional testing. *Alternative assessment* asks students to show what they can do; students are evaluated on what they integrate and produce rather than on what they are able to recall and reproduce (Huerta-Macias, 1995). Competency-based assessment demonstrates what students can actually *do* with English. Alternative assessment differs from traditional testing in that it:

- does not intrude on regular classroom activities
- reflects the curriculum actually being implemented in the classroom
- provides information on the strengths and weaknesses of each individual student
- provides multiple indices that can be used to gauge student progress
- is more multiculturally sensitive and free of the linguistic and cultural biases found in traditional testing (Huerta-Macias, 1995).

Types of Alternative Assessment

Several types of alternative assessment can be used with great success in today's language classrooms:

- Self-assessment
- Portfolio assessment
- Student-designed tests
- Learner-centered assessment
- Projects
- Presentations

Specific types of alternative assessment will be discussed in the skills chapters.

This chart summarizes common types of language assessment.

Table 1: Common Types of Language Assessment	
Informal	Formal
Classroom, "low-stakes"	Standardized, "high-stakes"
Criterion-referenced	Norm-referenced
Achievement	Proficiency
Direct	Indirect
Subjective	Objective
Formative	Summative
Alternative, authentic	Traditional tests

Because language performance depends heavily on the purpose for language use and the context in which it is used, it makes sense to provide students with assessment opportunities that reflect these practices. *Our assessment practices must reflect the importance of using language both in and out of the language classroom.*

It is also important to note that most testers today recommend that teachers use multiple measures assessment. Multiple measures assessment comes from the belief that no single measure of language assessment is enough to tell us all we

need to know about our students' language abilities. That is, we must employ a mixture of all the assessment types previously mentioned to obtain an accurate reading of our students' progress and level of language proficiency.

Test Purpose

One of the most important first tasks of any test writer is to determine the purpose of the test. Defining the purpose aids in selection of the right type of test. This table shows the purpose of many of the common test types.

Table 2: Common Test Types	
Test Type	Main Purpose
Placement tests	Place students at appropriate level of instruction within program
Diagnostic tests	Identify students' strengths and weaknesses for remediation <i>(rimidiyegun)</i>
Progress tests or in-course tasks	Provide information about mastery or difficulty with course materials
Achievement tests	Provide information about students' attainment of course outcomes at end of course or within the program
Standardized tests	Provide measure of students' proficiency using international benchmarks

Timing of the Test

Tests are commonly categorized by the point in the instructional period at which they occur. Aptitude, admissions, and general proficiency tests often take place before or outside of the program; placement and diagnostic tests often occur at the start of a program. Progress and achievement tests take place during the course of instruction and promotion, while mastery or certification tests occur at the end of a course of study or program.