

Course Topic: Language Assessment**Course Instructor:**

Julia Williams has been in the English language education sector for over 20 years. She has taught at universities, colleges and secondary schools in Ontario and is currently the director of English Language Studies and Coordinator of Applied Language Studies at Renison University College, University of Waterloo. She is the author of both the first and second editions of *Learning English for Academic Purposes: Reading and writing*; as well as books 3 and 4 in the *Academic Connections* series of textbooks published by Pearson. Julia has presented nationally and internationally on language assessment, English for Academic Purposes, occupation-specific language training and the effective use of textbooks.

Course Synopsis:

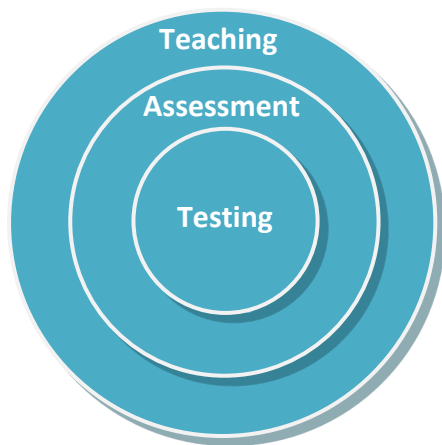
In this course, we will explore the principles of testing and assessment in an ESL setting. We will examine the complex relationship between teaching, assessment and testing within the classroom. We will then explore the 5 principles of testing and assessment, which include validity, reliability, authenticity, practicality and washback. Finally, we will take a look at the application of the 5 principles to various assessment scenarios.

Course Agenda:

1. The Relationship Between Teaching, Assessment, and Testing
2. The 5 Principles of Testing and Assessment
3. Applying the Principles

1. The Relationship Between Teaching, Assessment, and Testing

Testing and assessment are distinct terms that encompass specific meanings in an educational setting. Teaching relates to everything a teacher does within the classroom, including providing information, feedback and support for students. Testing involves a traditional setting that includes a predetermined time constraint, paper and pencil, oral or multiple choice contexts. Assessment is a broad category wherein a teacher forms a judgment of a student based upon observable behaviours or characteristics. The bullseye graphic below illustrates the suggested proportion of assessment and testing in relation to teaching. While assessment and testing are closely linked, teachers should spend most of their time teaching, less time conducting assessment and even less time administering tests.



Assessment and teaching are very carefully linked. Useful questions to consider when reflecting on your teaching practice are ‘can there be teaching without assessment?’, or ‘can there be assessment without teaching?’ These are rhetorical questions with no right or wrong answer. Some teachers feel that teaching cannot take place without assessment. They feel that when you teach and work with students you are inevitably going to observe things that will tell you how well a student is learning. It is important to gauge how well a student is doing in order to help them achieve learning goals. On the other hand, some teachers feel it is important for students to work in a nonjudgmental environment where they can make mistakes free from criticism. These teachers will not pass judgments about a student until testing is conducted. Both sides of the argument are reasonable and important to think about when teaching and working with students. What is your opinion?

2. The 5 Principles of Testing and Assessment

There are 5 commonly recognized principles that apply equally to testing and assessment:

1. Validity
2. Reliability
3. Authenticity
4. Practicality
5. Washback

Let's examine each of these principles in detail.

1. Validity

There are many definitions of validity. It is a complex idea that many researchers and writers have written about. See the following definitions below.

The extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment.

Gronlund, (1998)

The degree to which a test measures what it claims, or purports, to be measuring.

Brown (1996)

The extent to which [a test] measures what it is supposed to measure and nothing else.

Heaton (1988)

Content Validity

Validity is a multifaceted construct. When considering validity, we need to ask what a test should test. A test is supposed to measure the criteria stated in the objectives and the elements taught in class. A good match between what you have taught and what you are testing is referred to as content validity. Questionable content validity may occur in a final exam with a test of phonetics for a class taught primarily through oral tasks and assessments. A final exam that tests phonetics switches the skill set students learned in class and may not accurately reflect their learning skill development throughout the course.

Construct Validity

Content, organization and form are three key elements defined in second language research and in the teaching community as important subskills to the skill of proficient writing. Teachers must consider whether or not they are asking their students to write something where content, organization and form are tested. Form and accuracy refer to grammar, as it translates into spoken contexts when we look at the oral tests our students produce, which are related to pronunciation as well as accuracy in writing. Since content, organization and form are inextricably linked, we want to make sure we always provide marks for these skills separately. Teachers may add a mark for vocabulary range if it is a teaching objective that has been pursued in class. The question of construct validity recognizes if there is a match between what the test tests and the proficiency of a skill set determined by research and the teaching community. Questionable construct validity may occur in an essay writing class where students have produced different types of essays throughout the term, with a final essay writing exam that provides grades for organization and form and not for content. This exam constitutes questionable construct validity since it ignored one of the elements in essay writing that was a major focus throughout the term.

2. Reliability

Reliability affects testing and is related to the consistency of testing. It is multifaceted and can be examined in the context of student reliability, rater reliability and test reliability. A variety of factors can affect a student's state of mind before writing a test and may result in a score that does not

reflect true proficiency. For example, having an argument with a loved one, not feeling well, or temperature extremes in the room may create an unsettled frame of mind, making it difficult to concentrate. While it is hard for teachers to control certain factors, they can assist with issues relating to room temperature and background noise, while also displaying empathy and sympathy with student problems affecting testing abilities.

Rater Reliability

Rater reliability is related to how consistently the teacher scores the test. Factors including rater irritability, tiredness, or hunger can impact scoring by creating distractions and an unfocused frame of mind. Teachers may also harbor biases about particular students that may affect how well they score their tests, or they may forget to look for specific marking criteria. Rubrics allow teachers to maintain consistency when scoring an isolated skill across a wide range of tests. Teachers must ensure they do not rush through marking to compromise their reliability. *Intra-rater reliability* refers to the consistency of one scorer across multiple tests. *Inter-rater reliability* refers to the consistency of multiple scorers across multiple tests. Rubrics are used to maintain consistent scoring while benchmark sessions establish a common set of grading standards for scorers to adhere to before splitting up individually to score tests. Statistical analyses may be conducted to see whether or not consistency across multiple scorers is being achieved.

3. Authenticity

Authenticity relates to how close a match there is between the test task item and the types of things students will be required to do once they leave the classroom. An authentic test item is an activity that parallels what students might do in real-world situations. For example, testing grammar knowledge through multiple-choice is not an authentic test item since students are not likely to encounter multiple choice contexts in their real-world experiences. Teachers should therefore limit their use of multiple-choice and include other test items that require students to apply their knowledge to either speaking or writing. Testing language should also be authentic. Language should be as natural as possible and useful in contexts outside the classroom, where students produce answers that are natural in everyday conversation. Teachers should ensure the authenticity of context to match their teaching contexts with what is happening outside the classroom. An example of this would be teaching on a particular topic that students are likely to encounter in their daily lives.

4. Practicality

Practicality is important to consider when developing a test. Practicality relates to the logistics of the test itself, including the development of test tasks, scoring and providing timely feedback. Practicality also encompasses budgetary concerns regarding the cost of administering the test as well as the training of multiple scorers. For practicality reasons, a teacher may decide to test grammatical knowledge in the form of multiple-choice test tasks. While this may be a practical testing method, we must be sure to recognize that multiple-choice questions test only one element of grammatical accuracy. By acknowledging this limitation, a teacher can balance both principles of practicality and authenticity.

5. Washback

Washback refers to the concept of feedback. When developing a test, teachers must consider how they are going to provide feedback for students as well as the purpose of the test results. Are the test results intended for students, administrators, or for admissions purposes? Feedback can take many forms, including oral and written. Teachers must ensure they provide feedback in a positive, supportive and timely manner.

3. Applying the Principles

Now let’s take a look at how the principles apply to specific test examples.

Sample Test #1: Standardized Test of English Proficiency

- No oral skills are tested
- Writing is tested by multiple choice grammar questions
- Listening component is a poor quality sound file played in a large room
- Students receive a single score

	YES	NO
Valid? Construct validity: if writing is tested through multiple-choice tasks and there is no oral component, it is difficult to provide an accurate overall score for a student’s English proficiency.		✓
Reliable? Reliability may be variable throughout the test. Test reliability is quite low because the poor quality sound file will introduce elements of inconsistency for scoring.		✓
Authentic? Multiple-choice test tasks and the lack of an oral skills component mean the test is not authentic.		✓
Practical? The test is highly practical since it is inexpensive to administer, has objective scoring tasks, does not require working one-on-one with students and culminates in a single score.	✓	
Positive Washback? A single score does not provide positive washback, since students are not informed of their areas of strength and weakness.		✓

Sample Test #2: Mid-term Test based on Unit Objectives

- Students write an essay
- Teacher uses a rubric to score content, organization and form
- Test is returned within the week with individualized comments

	YES	NO
Valid? It is a valid test because there is a clear link between curriculum objectives and what students have learned in class.	✓	
Reliable? The rubric suggests the reliability of the test.	✓	
Authentic? Students need to write essays in academic contexts beyond the ESL classroom.	✓	
Practical? Rubrics ensure consistency; however, individualized comments are timely to provide for each student.		✓
Positive Washback? Individualized comments provide students with a positive overview of strengths, weaknesses and encouraging comments. Returning the essay within the week also ensures positive washback.	✓	

Thanks for participating!

www.aceducation.ca