EVALUACIÓN DEL APRENDIZAJE

EN INGLÉS

INDEX

Context

This course material has been created for the course "Evaluation of comprehension and oral and written production in English" (LCL 462) from the English Language Teaching (ELT) program at the Institute of Literature and Language Sciences (**http://www.ilcl.ucv.cl/**), Pontificia Universidad Católica de Valparaíso (PUCV) (**http://pucv.cl/**). This course is located in the 8th semester of the curriculum of the bachelor degree of "Pedagogy in English" (English Language Initial Teacher Training) at PUCV. This course corresponds to the area of specific disciplinary training, complements the professional training course "Evaluation of Learning" of the School of Pedagogy and precedes the professional practice, academic-professional activity culminating in the curriculum of this career.

This textbook compiles different sources of "open content" from other Open Educational Resources (OER), learning, teaching, and research materials that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions (UNESCO). Content of this textbook has been drawn from the United States Department of State website American English: For English Language Teachers around the World (**https://americanenglish.state.gov/**), whose content resides in the Public Domain (**https://www.state.gov/misc/87529.htm#copyright**).

The original content and activities, as well as the reused content has been built in this textbook have been selected, adapted and designed by the course lecturer, Jannett Fonseca PhD., with the help of Francisca Henríquez, senior student from the ELT program.

Contexto

Este material educativo de apoyo ha sido creado para el curso "Evaluación de la comprensión y producción oral y escrita en inglés" (LCL 462) del programa de Enseñanza del Idioma Inglés (ELT) en el Instituto de Literatura y Ciencias del Lenguaje (**http://www.ilcl.ucv.cl/**), de la Pontificia Universidad Católica de Valparaíso (PUCV) (**http://pucv.cl/**). Este curso se ubica en el octavo semestre del plan de estudios de la carrera de "Pedagogía en Inglés" (Formación inicial de profesores de inglés) en PUCV. Este curso corresponde al área de formación disciplinaria específica, complementa el curso de formación profesional "Evaluación de Aprendizaje" de la Escuela de Pedagogía y precede a la práctica profesional, actividad académica profesional que culmina el currículum de esta carrera.

Este libro de texto compila diferentes fuentes de "contenido abierto" de otros Recursos Educativos Abiertos (REA), materiales de aprendizaje, enseñanza e investigación que residen en el dominio público o han sido liberados bajo una licencia abierta que permite acceso, uso y adaptación sin costo. y redistribución por otros sin restricciones o limitadas (UNESCO). El contenido de este libro de texto ha sido extraído del sitio web del Departamento de Estado de los Estados Unidos, Inglés Americano: Para profesores de inglés de todo el mundo (**https://americanenglish.state.gov/**), cuyo contenido reside en el dominio público (**https://www.state.gov/misc/87529.htm#copyright**).

El contenido y las actividades originales, así como el contenido reutilizado que se ha creado en este libro de texto, han sido seleccionados, adaptados y diseñados por la docente del curso, Jannett Fonseca, PhD., con la ayuda de Francisca Henríquez, estudiante senior del programa ELT.

UNIT 1:
Exploring assessment principles

Learning Outcome

In this unit, you will become familiar with seven key concepts related to assessment, which sometimes are not made explicit in the planning stage. Not only will you read teacher-friendly definitions of those concepts but also you will be asked to apply those concepts when creating general test specifications for some of the courses in the ELT program.

Evidence of learning: Student creates exam specifications by collecting past exams, analysing test type questions and conferencing with ELT faculty.

Before we start!

📝 Task 1

1. Get together in small groups and discuss what you have learnt about assessment in your previous evaluation course (EPE 1302 Evaluación del y para el Aprendizaje).

2. Spend about 15 minutes sharing some key concepts, concepts that you learnt, or you are not sure about. Consider some key learning experiences from the course as well.

3. Create a poster to illustrate your knowledge and understanding (or lack of clarity). Be prepared to talk through your poster.

📝 Task 2

What does it take to be a teacher?

Watch the TedTalk (click on the link below) entitled 'Every kid needs a champion'. What is the call as a teacher? Does it have to do with knowledge, skills or dispositions? Discuss.
https://www.youtube.com/watch?v=SFnMTHhKdkw

Testing versus Assesing

The word test can make people nervous. It has semantic qualities that make us think of being judged or measured by someone or something. Many people have an emotive reaction to testing and associate it with negative experiences that they may have had as students. In an educational context, the terms testing and assessment are often used interchangeably to indicate the measurement of student learning. However, although a test is a type of assessment—usually thought of in the traditional sense of an exam or quiz—assessment is a more comprehensive term. It often indicates the collection of information about student learning that might include not only tests but also a variety of techniques such as performance tasks, portfolios, and observation.

While tests are thought of as a means to give grades to students, assessments offer diagnostic information for both students and teachers. The ultimate purpose of assessment is to improve student learning, as opposed to just being able to give a mark for the amount of course content a student has mastered. Today teachers tend to talk about assessing (rather than testing) their students because we see the ongoing evaluation of student learning as more than just testing knowledge and skills in a particular area at one point in time for grading purposes. Thus, throughout this article, references to tests will be made with the ultimate goal of using them as assessment tools and not purely as testing instruments.

Task 3

What is my personal experience with testing and assessment?

Procedure: Think back to what you have done both, when you were in high school and at the University and complete the table below.

Extracted from: Rogier, 2014 **https://americanenglish.state.gov/files/ae/resource_files/etf_52_3_02-13.pdf**

In high school (English classes)			
Type of assessment (make a list of assessment procedures you were asked to do)	What was being assessed? (grammar knowledge, speaking, reading, writing,listening, creativity?	Was there a blueprint?	Your experience (How did you feel about it/them?; Did you have to study a lot or very little? What were the stakes?
Ex: Questions/ answers with the teacher.	Mainly listening (understanding the input) and speaking (producing the expected answer).	No. We became familiar with the procedure because of practice.	It was very stressing. Teacher would call sts randomly. The student had to answer the teacher's questions in front of the whole class. There was no escape.

At the University (English classes)			
Type of assessment (make a list of assessment procedures you have been asked to do)	What is being assessed? (grammar knowledge, speaking, reading, writing,listening, creativity?	Is there a blueprint?	Your experience (How do you feel about it/them?; Do you have to study a lot or very little? What are the stakes?
Ex: Dramatizations	Mainly speaking (from a script written beforehand)	Some general guidelines only.	It took quite a lot to prepare but it was fun. I particularly enjoy role-playing so it feels kind of natural.

Share your notes and reflections with the class. Reflect on what testing and assessment have meant to you as a class.

Importance of Assessment

As we all know, assessment plays an important role in teaching and learning. It affects decisions related to instruction, determines the extent to which instructional objectives are met, and provides information for administrative decisions. It has been estimated that teachers spend as much as 50 percent of their time in assessment-related activities (Stiggins 1991), and that when assessment is implemented effectively, student achievement is improved (Campbell and Collins 2007).

Yet many teachers feel assessment and testing are not relevant to their classroom practice and report that they feel unprepared to undertake assessment-related activities. Popham (2004) reports that most public school educators in the United States tend to think of assessment as "a complex, quantitative arena well beyond the comprehension of mere mortals". Some of these feelings may come from the anxiety that teachers felt when they were students taking tests, especially if they didn't understand how the tests were graded or if the objectives of the tests weren't clear. Teacher-education programs are also at fault for not making sure teachers are adequately trained before entering the classroom (Mertler 2004). As Taylor (2009) points out, language education programs at graduate level typically devote little time or attention to assessment theory and practice, perhaps just a short (often optional) module; and although there is no shortage of books on language testing and assessment available today, many of these are perceived to be (and often are) highly technical or too specialized for language educators seeking to understand basic principles and practice in assessment.

During our time in school and teacher-training courses, we take many tests, but how often are we actually given practice creating them, marking them, and interpreting the results? Developing these skills is part of becoming assessment literate.

Assessment Literacy

Task 4

How assessment literate are you?

Procedure:

a) Complete the checklist below.

b) Analyse your answers within a small group of classmates. What can you infer from your personal experience and that of others' in your group.

c) Save this checklist. You will be asked to go back to it towards the end of term.

How assessment literate are you? Tick the columns that best represents your experience.

Assessment experience	Never	Seldom	Sometimes	Frequently
I have created quizzes for school students				
I have created quizzes for classmates				
I have created quizzes in collaboration with others.				
I have created summative tests for school students.				
I have checked and marked students' tests.				
I have given oral and/or written feedback to students.				
I have given oral and/or written feedback to classmates				
I have created rubrics basically using other people's rubrics.				
I have created rubrics in collaborations with others.				
I have experienced teachers analyzing partial and global results using statistics.				
I have analysed partial and global results using statistics.				
I have read the guidelines for assessment as specified by the MINEDUC.				
I have made enquiries into assessment for all (inclusion)				
I have created accommodations for assessment procedure.				

An essential element of assessment literacy is the ability to connect student assessment to the learning and teaching process. Teachers can make this link by first matching test items to instructional objectives, then using the test results to provide feedback on both student performance and how well the instructional objectives were met. An assessment-literate teacher is able to interpret data generated from a test to make useful modifications to teaching and to use assessments as a tool to improve student learning. Assessment-literate teachers are also able to discuss assessments with others in terms of key concepts in testing. With this in mind, we can explore common terms associated with tests, along with their practical application.

Key concepts and considerations

Seven key concepts—usefulness, reliability, validity, practicality, washback, authenticity, and transparency—are cornerstones in testing that help to ensure that a test is solid (i.e., that it will consistently measure what you want it to measure in an efficient manner, and that both teacher and student will see it as a valuable source of information regarding learning). Understanding these concepts and being able to improve practices related to them are important in developing assessment literacy. Each is discussed separately below, but as you will notice, they are connected to and support one another; together, they form the basis for building solid assessments.

1.   Usefulness and purpose

According to Bachman and Palmer (1996), usefulness is the most important consideration when choosing or designing a test. Teachers must consider what the purpose of a particular assessment is and whether this purpose is congruent with the students they are testing and the course they are teaching. All language tests must be developed with a specific purpose, a particular group of test takers, and a specific language use in mind. Even tests with the general purpose of testing English language ability (proficiency) are designed with a specific group of test takers in mind. Take, for example, three standardized tests used globally for the purpose of measuring language ability: the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), and the Michigan English Test (MET).

Tests are designed with very specific audiences and purposes in mind. This specificity is what allows them to effectively measure what they are designed to measure and makes them useful for a specific purpose. You must carefully consider the purpose of a test before administering it. If you choose a pre-made test and it does not match your students' needs or your purpose, then it will not be an adequate assessment of your students and will not provide the information that you need in order to make informed decisions about the teaching and learning taking place in the classroom.

For example, if you wanted to measure the reading ability of your students to see if they would be able to order from a menu when visiting the United States on an exchange trip, you couldn't just use any reading test you find in a textbook or online. You would need to find one (or better yet, make one) that is specific to the skills taught in class, that meets the vocabulary needs of the situation the students would be immersed in, and that uses an appropriate text style that matches what you expect the students to encounter. Having them read a passage from a newspaper or a short story and then answer questions would not adequately measure their ability to read and order from a menu at a restaurant. So when you choose or design a test, consider the purpose of the test, the group of test takers it is designed for, and the specific language use you want to evaluate.

2. Reliability

Your assessments not only need to be useful for the intended purpose, they also need to be reliable. Reliability refers to the consistency of test scores. If you were to test a student more than once using the same test, the results should be the same, assuming that nothing else had changed. Reliability can be threatened by fluctuations in the learner, in scoring, or in test administration. Fluctuations in the learner are out of the testing administrator's control; we cannot control whether a student is sick, tired, or under emotional stress at the time of a test. But we as teachers can limit the fluctuations in scoring and test administration. The guidelines for how a test is administered, the length of time allotted to complete the test, and the conditions for testing should be established in advance and written in a test-specifications document.

As much as possible, there should be consistency in testing conditions and in how a test is administered each time it is given. Teachers can minimize fluctuations in score by preparing answer keys and scoring rubrics, and by holding norming sessions with those who will be scoring the test.

You can take steps to improve the reliability of your tests. You need to make sure that the test is long enough to sample the content that students are being tested on and that there is enough time for most of the students to finish taking the test. The items should not be too easy or too difficult, the questions should not be tricky or ambiguous, the directions should be clear, and the score range should be wide. Before you administer the test, you might want to have someone else take it to see whether he or she encounters problems with directions or content. Use that person's feedback to see where the test might need to be improved.

3. Validity

One thing to keep in mind is that a test may be highly reliable, but not valid. That is, it might produce similar scores consistently, but that does not mean it is measuring what you would like it to. A test has validity when it measures what you want it to measure. The most important aspect of validity is the appropriateness for the context and the audience of the test. Think about what is to be gained by administering a test and how the information will be used. Suppose your goal is to measure students' listening ability, and you give a test in which students answer questions in written format about a lecture they hear. In that case, you need to make sure that the vocabulary, sentence structure, and grammar usage in the written questions are not beyond the level of the students. Otherwise, you will be testing them on more than just their listening comprehension skills and thus decreasing the validity of the test as a measure for listening ability.

A number of factors can have an adverse effect on validity, including the following:

- unclear directions

- test items that ask students to perform at a skill level that is not part of the course objectives

- test items that are poorly written

- test length that doesn't allow for adequate sampling or coverage of content

- complexity and subjectivity of scoring that may inaccurately rank some students

The best way to ensure validity and reliability is to create test specifications and exam blueprints. These will help ensure that tests created and used match what is intended for the course and the students. Figure 2 shows an example of

general information for the test specifications of a final exam for a higher-education pre-academic English-language program course. For each of the subtests (listening, reading, and writing), specifications would also be written and would include the type of skills being assessed, level of vocabulary, grammar structure, and length of text to be used.

4.   Practicality

Practicality refers to how "teacher friendly" a given test is. Practicality issues include the cost of test development and maintenance, time needed to administer and mark the test, ease of marking, availability of suitably trained markers, and administration logistics. If the test you want to give requires computers, and these are not available or connectivity is unreliable, there will obviously be a practicality issue with the delivery of the test. For many teachers, the amount of time required to mark a test is an important practicality issue. You can overcome this issue by weighing how important a particular assessment is in terms of overall course mark and determining how much time you want to spend marking it. For example, if a vocabulary quiz will not be worth much in the overall course mark, you might consider having students exchange papers and mark them instead of marking each one yourself. This arrangement also allows students to review the materials at the same time. For marking writing, it might be more practical to have students review each other's work and peer edit the first draft than to have the teacher make comments on each initial draft.

5.   Washback

Washback refers to the effects of testing on students, teachers, and the overall program. It can be positive or negative. Positive washback occurs most often when testing and curriculum design are based on clear course outcomes that are known to all students and teachers. On the other hand, exams that require extensive preparation can have negative washback and be harmful to the teaching and learning process; if instruction solely focuses on helping students pass the test, other learning activities may be neglected. To make sure washback is positive, teachers should link teaching and testing to instructional objectives. Tests should reflect the goals and objectives of the course along with the types of activities used to teach the content. That underscores the importance of planning assessments at the same time you plan the course.

Another way to bring about positive washback is through feedback. Providing feedback in a timely manner is important if you want students to learn and benefit from the assessment process. In the above example of practicality, having students mark their classmates' papers provides timely feedback to the students and helps them understand where they might need further practice or review. Using short quizzes that are graded immediately by the students throughout the course may let students know where they need to study more; it may also redirect teacher energy toward the areas that need more instruction time.

Involving students in the marking process is one way to create positive washback from testing. Other ways are to use authentic testing materials and to make the assessment procedures transparent—the topics of the next two sections.

6.   Authenticity

Tasks that reflect real-world situations and contexts in which the language will be used provide motivation for learners to perform well. Assessment tasks should be relevant to real-life contexts in which the language will be used. For example, if a course is designed for students who will be answering phones in English in a call center, an oral exam that mimics a telephone-call format would be more authentic than a test in which students listen to an academic lecture and respond to questions related to the lecture, or one where the students write the correct forms of verbs in sentence blanks. The assessments should relate to the purpose of the course, which in turn relates to course objectives, which are then tested on the assessments.

7.   Transparency

Transparency refers to the availability of information to students. Students should be aware of the skills, vocabulary, and grammar that they will be expected to learn, and they should receive a clear explanation of how these will be assessed. Transparency makes students part of the testing process by ensuring that they understand what the course objectives are and what will be tested, as well as the format of tests and how they will be used and graded. Students should have the chance beforehand to practice question types that will be used in a test. Using a new test format, one that students are unfamiliar with, could affect the test's reliability. When students do not perform well on a test, it should be because they have not learned the material, not because they didn't understand the directions to complete a task.

Increasing transparency will also reduce students' test anxiety and allow them the chance to perform better. To increase transparency, many schools and educational institutions publish their test specifications. For example, the Oregon Department of Education publishes test specifications for the English Language Proficiency Assessments by grade level on its website (Oregon Department of Education 2014). These documents list not only content to be tested, but also in what ratio, along with appropriate test-item types.

Figure 1. Example of general test specifications

General Test Information — Final Exam

Purpose (Why are you testing?)	To test student mastery of listening, reading, and writing curricular objectives for Level 1
Intended population (Who are you testing?)	Students in university pre-academic intensive English program – Level 1
Intended decisions/stakes (How important is the test for the course grade?)	High stakes – weighted as 40% of the final grade
Response format (What type of questions will you use? How will the test taker show mastery of the objective?)	Listening: multiple choice, short answer, matching, gap fill, and information transfer
	Reading: multiple choice, short answer, matching, gap fill, and information transfer
	Writing: one-paragraph response to a prompt (input is a picture or personal knowledge)
Number of examiners (How many people are needed to administer the test? Are there any restrictions for test supervisors?)	One test supervisor per 20 students; two markers per exam (cannot be the class teacher)
Number and weighting of items/ tasks (How many questions will there be on each part? How much will each part be worth for the overall grade of the test?)	Listening: approximately 20 items (33%) Reading: approximately 20 items (33%) Writing: 1 task (34 %)
Examination length (How much time will the assessment take overall? Is there a time length per section?)	Maximum of 2 hours Listening: 30 minutes Reading: 40 minutes Writing: 40 minutes
Order of tasks (In what order will the sections be tested?)	1. Listening 2. Reading 3. Writing
Rating scale type (Conditions necessary for marking the exam)	Reading and Listening: Answer key agreed to before the test; Writing: Two markers, analytical criteria, third marker if necessary
Reporting type (How will the score be reported? As a whole score, or per section? What is the passing grade?)	Single score (maximum 100%; pass mark 70%)

Now that you are familiar with the seven cornerstones of assessment, let's examine how you would go about planning an assessment that is useful, valid, reliable, practical, authentic, and transparent, and that has positive washback. Planning your assessments goes hand in hand with developing your course learning objectives and should start when you begin planning the course. How you will assess student learning will affect how you present materials and teach the course. There are several phases in the assessment process. One of the most important is the initial planning stage. When you plan an exam, begin by describing your assessment context. Think about what the purpose of the course is, which resources you have available, and how the instructional setting and larger educational context influence the course. This is the information that you will put in the test specifications, discussed above in the Validity section, in the categories for purpose and intended population.

Task 5

General Test Specifications - Final Exams for face-to-face classes

Procedure:

A.   Use the template above as a guideline to create the General Exam Information for ONE of the following courses: LCL121, LCL135, LCL216, LCL263, LCL303 and LCL 338, LCL463 and LCL547.

B.   Use the template above to create the General Test Information for the final exams for either a literature, linguistics and/or didactics course.

C.   State the validation procedures you have carried out to ensure the GTI is accurate and reliable. For example: have you conferenced with colleagues?; have you examined past exam papers?; have you checked the information against the course syllabus?

D.   Reflection log: Write an entry (one-two page long) in your reflection log. State what you have learnt by carrying out this activity both as a student as well as a would-be teacher. For your analysis, you might consider the conversation with your teachers, the decision-making mechanisms, teachers' metacognition, amongst others.

E.   Submit your work to the Learning Management System.

F.   Discuss your work with your classmates in a 10 minute presentation format.

General Test Specifications - Final Exams for online or blended courses.

G.   Use the template above as a guideline to create the General Exam Information for an English course taught at your local institution.

H.   Use the template above to create the General Test Information for the final exams for either a literature, linguistics and/or didactics course at your local institution.

I.   State the validation procedures you have carried out to ensure the GTI is accurate and reliable. For example: have you conferenced with colleagues?; have you examined past exam papers?; have you checked the information against the course syllabus?

J.   Reflection log: Write an entry (one-two page long) in your reflection log. State what you have learnt by carrying out this activity both as a student as well as a would-be teacher. For your analysis, you might consider the conversation with your teachers, the decision-making mechanisms, teachers' metacognition, amongst others.

K. Create a 5 to 10 minute screencast video showcasing the information you have collected and the completed GTI form. As you go over the different sections, share how you collected the information, what was discussed with the class teacher whose exam you are creating the GTI for, insights into the decision-making process of the teacher(s) and the institution. You may use Jing, ScreencastOMatic, Camtasia or others.

>   **https://www.techsmith.com/jing-tool.html**
>   **https://screencast-o-matic.com/**
>   **https://www.techsmith.com/video-editor.html**

L. Submit your work on the Learning Management System.

General Test Information — Final Exam	
Purpose (Why are you testing?)	
Intended population (Who are you testing?)	
Intended decisions/stakes (How important is the test for the course grade?)	
Response format (What type of questions will you use? How will the test taker show mastery of the objective?)	
Number of examiners (How many people are needed to administer the test? Are there any restrictions for test supervisors?)	
Number and weighting of items/ tasks (How many questions will there be on each part? How much will each part be worth for the overall grade of the test?)	
Examination length (How much time will the assessment take overall? Is there a time length per section?)	
Order of tasks (In what order will the sections be tested?)	
Rating scale type (Conditions necessary for marking the exam)	
Reporting type (How will the score be reported? As a whole score, or per section? What is the passing grade?)	

Closing: Summary of key concepts

(Source: Dendrinos, Bessie (2015), "ELT Methods and Practices. Testing and assessment", Copyright National and Kapodistrian University of Athens. Edition: 1.0. Athens 2015. Available at: http://opencourses.uoa.gr/courses/ENL4)

Validity is commonly defined as 'the extent to which [a test] measures what it is supposed to measure and nothing else. If a test is valid, the outsider who looks at an individual's score knows that it is a true reflection of the individual's skill in the area the test claims to have covered.

- Content validity: A test is said to have content validity if the items or tasks of which it is made up constitute a representative sample of items or tasks for the area of knowledge or ability to be tested (often related to a syllabus or a course).

- Construct validity: A test is said to have construct validity if the scores that a candidate gets on this relate in the same way to another test or form of assessment for the same aspect of knowledge.

- Empirical validity: A measure of the validity of a test arrived at by comparing the test with one or more criterion measures.

- Face validity: The extent to which a test appeals to candidates or to those choosing it on behalf of the candidates because it is considered to be an acceptable measure of the ability they wish to measure. It is sometimes referred to as 'test appeal'.

- Predictive validity: A type of validity based on the degree to which a test accurately predicts future performance. A language aptitude test for example, should have predictive validity because the results of the test should predict the ability to learn a foreign language.

Reliability is another very important consideration when testing.

- Reliability: refers to the consistency of a test. That is, if every time the test is administered it will have the same outcome. But reliability does not have to do with the content of the test alone; it has to do with marking in two ways:
  - ensuring that different raters give comparable marks to the same script,
  - the same raters give the same marks on two different occasions to the same script.

- Reliability is most often estimated with regard to:
  - The internal consistency in a test; that is, if there is correlation among the variables comprising the test.
  - The results when testing and re-testing; that is, if there is correlation between two (or more) administrations of the same item, scale, or instrument for different times, locations, or populations, when the two administrations do not differ in other relevant variables.

- Inter-rater reliability, which refers to the level of agreement between two or more evaluators/ judges/ raters on a particular instrument at a particular time. They are to apply their marks in a manner that is predictable and replicable. Therefore, note that inter-rater reliability is a property of the testing situation, and not of the instrument itself.

END OF UNIT 1

UNIT 2:
Assessing receptive skills

Learning Outcome

In this unit you are expected to demonstrate understanding in planning, writing, applying and reflecting on different assessment tools, both for reading and listening skills. The planning stage includes writing the test blueprints; the writing stage includes the alignment of the question types with the learning objectives set for the course. You are expected to apply the test to a group of students, analyse the results quantitatively and reflect on what the results tell you about the test itself, the questions and students' learning. You are expected to apply the key assessment concepts from Unit 1.

Evidence of learning: Student creates a set of reliable and validated reading and listening comprehension tests. Student applies descriptive statistical analysis.

Before we start!

📋 Task 6

Get together in small groups and jot down some ideas to the following questions:

1. As compared to reading a written text, how is listening different or similar?

2. What strategies do L2 learners -yourselves- use when practising listening comprehension exercises?

3. Based in your own experience as language learners, provide a working definition for the following concepts. Add an example of question format for each.

a) Intensive listening
b) selective listening
c) extensive listening
d) extensive reading
e) elective reading

4. Plenary: In a plenary format, share your notes with the rest of the students in your class.

Assessing reading: An overview

Brown (2004) has identified 10 key strategies for reading comprehension, which are organised taxonomically as follows. Readers:

1. identify their purpose for reading a text

2. appy spelling rules to aid bottom-up decoding

3. use lexical analysis to determine meaning (suffixes, prefixes, roots)

4. infer the meaning of words, idioms and expressions they are not certain of

5. skim the text for gist and main ideas

6. scan the text for specific information (names, places, dates, prices, keywords)

7. use silent reading techniques for rapid processing

8. use notes on the margin, outlines, charts, or semantic maps for understanding or retaining information

9. distinguish between literal and implied meanings

10. capitalize on discourse markers to process coherence and cohesion.

Types of reading

Brown (2001, 2004) states that a variety of performance is derived from the multiplicity of genres of written texts. However, for assessment procedures, some reading performance serve as organizers of various assessment tasks. These types of reading are:

- **Perceptive:** Perceptive reading tasks aim at attending to the components of larger stretches of discourse: letters, words, punctuation and other graphemic symbols. Bottom-up processing is implied.

- **Selective:** This is typically an artifact of assessment formats, for example: picture-cued tasks, matching, true/false, multiple choice, among other close-ended and brief-responses tasks. A combination of bottom-up and top-down processing may be used.

- **Interactive:** It includes longer texts (paragraphs and one-page long texts). in which the reader must 'interact' with the text, in other words, the reader must 'negotiate meaning' by bringing to the text a set of schemata (knowledge of the world, knowledge of the language, knowledge of the particular genre). Typical genres for this type of reading are anecdotes, short narratives, news coverage, editorials. Top-down processing is typical of such tasks, and only some bottom-up processing.

- **Extensive:** It applies to texts of more than a page, up to including professional articles, reports, essays, short stories and books. The purposes of assessment usually are to tap into a learner's global understanding. Top-down processing is pivotal for most extensive tasks.

Assessing listening: An overview

Just as there are different types of reading, it is also possible to identify types of listening. Successful listeners, just as successful readers, are good at selecting the type of listening that suits the kind of input they encounter (Green, 2014) and the aims they pursue. Field (2008) stated that most listening assessments involve a relatively narrow range of these possible listening types, leaving out real world listening experiences (face-to-face interaction with other people).

Types of listening

As for reading, Brown (2004) identifies four type of listening performance. These are:

- **Intensive:** Identification of the components (phonemes, words, intonation, discourse markers, among others) of a larger stretch of language.

- **Responsive:** Listening to a relatively short stretch of language in order respond in equally short terms (responding to a greetings, questions, commands).

- **Selective:** For selective listening, learners listen to fairly short monologues and dialogues (up to several minutes) in order to select specific information. The focus is not on understanding global or general meanings. Typically, assessment tasks require learners to listen for names, numbers, directions, facts and events.

- **Extensive:** Listening to develop a top-down, global understanding of spoken language. It usually includes lengthy lectures and interviews. Listening for gist, for main ideas and making inferences are all part of extensive listening.

Why pilot tests?

As discussed in Unit 1, practicality, validity, reliability do not just happen. Teachers have to ensure that the instruments they design meet benchmarks, are valid and reliable in more than one aspect and, meet practicality criteria as well. This is especially crucial in achievement assessment and high-stakes testing.

We will now focus our attention on designing valid and reliable tests to assess reading and listening comprehension in the L2 classroom. This will be done by carrying out two piloting procedures, one to assess reading and the other to assess listening comprehension.

Hence, the first step is to identify students' needs and develop course learning objectives. Learning objectives are determined by what you want your students to know—and may be mandated by institutional or national priorities for education within your context. You should specify what you want your students to learn or be able to do upon completing a content unit or after taking the course. This will guide you in developing not only lessons and curriculum, but also in deciding how you will assess whether students have learned what you want them to. Identifying course learning objectives will give you and the students goals to work toward during the course.

Each of these objectives can then be divided into the skills needed to accomplish the objectives, whether they relate to vocabulary, structure, or fluency skills. With these learning objectives in hand, you will be able to design a test and check that the test you hope to use will accurately measure these objectives. The best way to do this is to create a blueprint of the assessment, matching course objectives to the test questions. By using the course learning objectives to guide the content and the purpose of your exam, you can make sure that your assessments serve both as a tool for providing information about student learning and as a means of assessing the course materials and instructional practices.

Having an exam blueprint increases the likelihood that you will actually test what you set out to test (i.e., the test will have validity). Test blueprints help you avoid overemphasizing one area or completely missing another area that needs to be tested. A blueprint is a tool to determine what is important for the students to know and the relative weight of each area in relation to other areas or skills being tested; at the same time, a blueprint ensures that the content being taught is properly represented on the test. The blueprint can also help a teacher see that the method used for assessing matches the cognitive demand that is intended. Begin creating a blueprint by listing

the learning objectives you want to measure, the way they will be tested, and how much of the total exam will cover each area. There may be several items on the exam related to each objective, but by first mapping out what you hope to test, you can be sure to include questions that assess all your objectives. An example of a simplified blueprint is given in Figure 2.

It lists the skill area (in this case, the skills to be tested are reading, listening, writing, and grammar), the learning objective, the item/question type, and overall percentage of importance in the context of this assessment. You can also select a test that is already made and map it backward to see if it will fit your purposes or if items need to be added, adjusted, or replaced. To map backward, you would list each question, what it tests, and the number of points it is worth. At the end of this exercise, you should be able to see what content is being tested and whether it is tested in the correct proportion to what you hope the students are learning.

Figure 2. Simplified test / exam blueprint

Skill	Learning Objective	How Objective(s) Will Be Tested	Total
Reading	• Can scan to find specific information • Can recognize main idea of a paragraph • Can understand pronoun references	Read a paragraph and answer questions related to a reading passage using multiple-choice, short-answer, matching, gap-fill, and information-transfer items	30%
Listening	• Can recognize main idea of a section • Can listen for specific information • Can listen for numbers	Listen and answer questions using multiple-choice, short-answer, matching, gap-fill, and information-transfer items	30%
Writing	Can use pronouns to show cohesion between sentences	Write sentences related to a personal topic	15%
Grammar	• Present simple • Adjectives • Subject/object pronouns	Multiple-choice and fill-in-the blank items	25%
Total			**100%**

Writing questions

As previously discussed, assessment processes involve different stages, including writing questions. It is by far the most critical element as it is through questions -or tasks if you are working on a task-based syllabus- that tests can be reviewed against validity and reliability. Questions or tasks must be aligned to the course learning outcomes and the construct and content validity of the instrument.

Either as a novice or experienced teacher, it is advisable to refer to a taxonomy. First, it guides the complexity level of the questions or tasks; secondly, it provides samples of questions for each level or category. Below is the Revised Bloom's Taxonomy, one of the most referred taxonomies in education. Fig. 3.
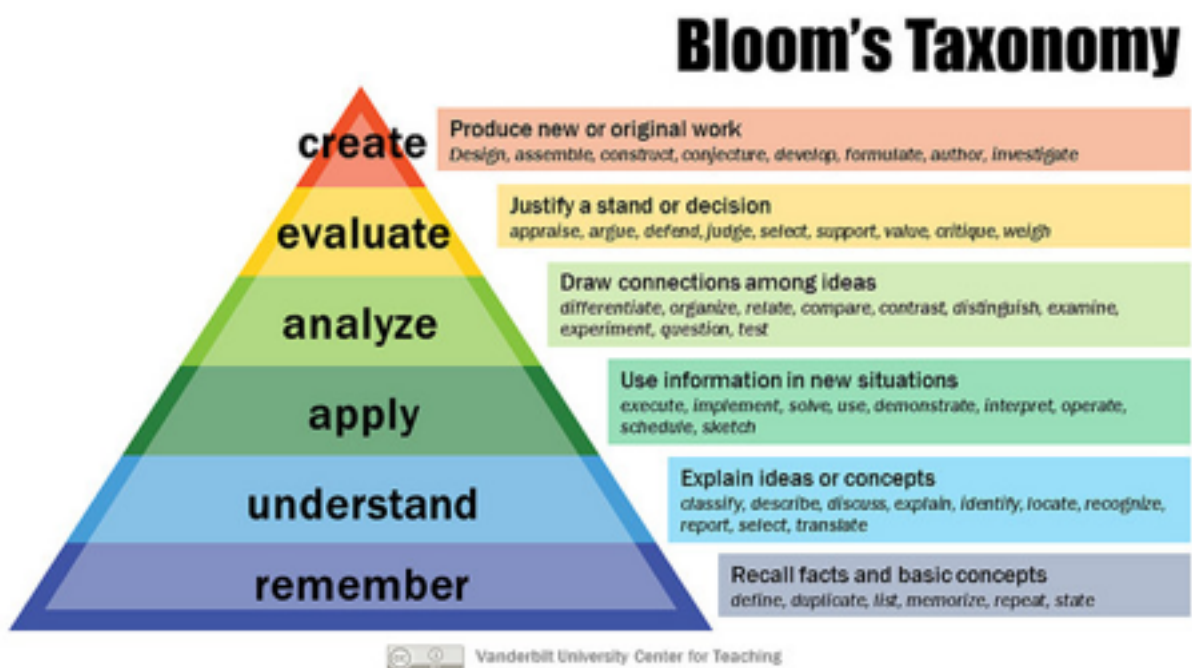

Fig. 3 The above graphic is released under a Creative Commons Attribution license. You're free to share, reproduce, or otherwise use it, as long as you attribute it to the Vanderbilt University Center for Teaching.

Figure 4 shows sample questions that might be on a reading test, in this case, the topic of the passage was dogs.

Figure 4. Analysis of exam questions relative to learning objectives

Item	Learning Objective	Item Type

What is the main idea of the reading?
    A. How to care for a dog
    B. The many different breeds of dogs
    C. The many ways that dogs are important to people
	Reading: Can understand main idea	Multiple-choice question

In Paragraph 1, what does they refer to?
    A. veterinarians
    B. dog trainers
    C. dogs
	Reading: Can understand pronoun references	Multiple-choice question

What is the meaning of the word flush in Paragraph 3?
    A. raise
    B. remove
    C. even with
	Reading: Can understand vocabulary in context	Multiple-choice question

What is the main idea of Paragraph 3?
    A. Dog showing is a popular sport.
    B. There are several hundred breeds of dogs.
    C. Obedience training for dogs is important.
	Reading: Can recognize main idea of a paragraph	Multiple-choice question

According to the reading, which is considered a sporting dog?
    A. collie
    B. fox terrier
    C. pointer
	Reading: Can scan to find specific information	Multiple-choice question

Match the type of dog with its description, according to the information in the reading.
 	
basset hound _____
poodle _____
Chihuahua _____
terrier _____
 	
    A. is trained to pull sleds
    B. has long ears
    C. is the smallest pure-bred dog
    D. sheds very little hair
    E. is used to herd animals
	Reading: Can scan to find specific information	Matching

The questions have been analyzed to determine the objectives being assessed and the mix of item types being used. One way to develop items for a test is to write them on notecards (or if you have a computer, add them to a spreadsheet) so that you can then sort them according to skill, question type, or objective covered. This system will let you know whether you have too many questions of the same type or need to add more in a certain category. Writing questions at the end of each lesson you teach is also an effective way to reflect on what you taught during the lesson and how you could assess that content in a manner the students will be familiar with. Working with a group of teachers to make new test items, or even working with your students, is a way to add variety to your item bank (your collection of possible test questions) and to get other opinions on what might be appropriate.

Having students become part of the assessment process can increase transparency, help them recognize what is important to know from the content being taught, and motivate them to review and study together. If students have a clear understanding of what they are being tested on and how they will be tested, positive washback is promoted. If you are working at an institution that uses very high stakes testing and standardized exams for all students, you might consider having the students create assessments for review or quizzes. For example, you could ask each student to write questions for a particular grammar item or vocabulary list. This method will also help you informally assess student knowledge of the topics being taught as you review students' questions and answers to decide whether they are suitable for a small-scale assessment.

After questions are written, label each with the learning objective that it covers and the item type. You can then look at the questions, make selections, and put them together for the quiz. You can look at item cards your - self or with other teachers to determine the strength of each question for testing a given objective. Looking at questions with others is a useful practice that can also generate more items or improve the ones you have. You can make stacks of questions relating to the same curriculum objective and determine how many of each you will need for the appropriate weighting on the test or quiz. Once the assessment has been administered, you will also want to add information to each question about how it performed (see Basic Statistics for Testing, Unit 3). These testing or item statistics will be useful in determining whether to use the questions again and which learning objectives might need to be reviewed or taught in a different manner in future lessons. In this way, an assessment can provide both teacher and student feedback.

Figure 5 below showcases stem questions and verbs associated to Bloom's Taxonomy. It is advisable to use it as reference when designing questions and tasks both for formative and achievement assessment.

Bloom's Taxonomy: Teacher Planning Kit

Source: **https://www.virtuallibrary.info/blooms-taxonomy.html**

Task 7

Procedure:

Before the meeting with the students, go over the check-list below:

Check list	Yes	No
I contacted students via email		
I confirmed date and place via email or via a social media app		
I have written a learning objective for the test		
I have awarded points to each section of the test		
I printed the necessary number of copies (text and questions, survey)		
I know exactly what I need to tell / explain to the students		
I created the answersheet of the test		

Note: Add rows if necessary

Data collection.

a) Time the test

b) Take notes on the experience itself: Do the students read the text once, twice? Do they scan for information? Do they ask questions? (Can they ask questions?); Do they answer straight away without using the text after reading it for the first time?, etc.

c) Ask follow up questions: Was the text difficult, easy for you? Why so? Any particular problems with the questions?

Analysis

Consider the following points for your analysis. You can add others as well.

- Content validity
- Construct validity
- Face validity
- Item reliability
- Level of difficulty of the text and questions; target audience for this test (A1- B2 level students? High school students?

Submit your assignment to the LMS.

Suggested source for reading materials **https://americanenglish.state.gov/files/ae/resource_files/03_max_6-7.pdf**

NOTE: The same task and procedure may be replicated for designing and piloting listening assessment. Instructors and/or students may opt to do both or choose either reading or listening assessments.

Basic Statistics for Assessment

Source:**https://cnx.org/contents/MBiUQmmY@19.6:y0GFmfab@8/Definitions-of-Statistics-Prob** This work is licensed by OpenStax under a Creative Commons Attribution License (by 4.0).

Task 8

Collaborative Exercise:

a) In your classroom, try this exercise. Students from a class have written down down the average time (in hours, to the nearest half-hour) they sleep per night. For example, consider the following data:

5 5.5 6 6 6 6.5 6.5 6.5 6.5 7 7 8 8 9

b) Create a simple graph (called a **dot plot**) of the data. A dot plot consists of a number line and dots (or points) positioned above the number line.

c) If you did the same example in an different class with the same number of students, do you think the results would be the same? Why or why not?

d) Where do your data appear to cluster? How might you interpret the clustering?

The questions above ask you to analyze and interpret your data. With this example, you have begun your study of statistics.

The science of **statistics** deals with the collection, analysis, interpretation, and presentation of **data**. We see and use data in our everyday lives.

Organizing and summarizing data is called **descriptive statistics**. Two ways to summarize data are by graphing and by using numbers (for example, finding an average). After you have studied probability and probability distributions, you will use formal methods for drawing conclusions from "good" data. The formal methods are called **inferential statistics**. Statistical inference uses probability to determine how confident we can be that our conclusions are correct.

Effective interpretation of data (inference) is based on good procedures for producing data and thoughtful examination of the data. You will encounter what will seem to be too many mathematical formulas for interpreting data. The goal of statistics is not to perform numerous calculations using the formulas, but to gain an understanding of your data. The calculations can be done using a calculator or a computer. The understanding must come from you. If you can thoroughly grasp the basics of statistics, you can be more confident in the decisions you make in life.

Key Terms

In statistics, we generally want to study a **population**. You can think of a population as a collection of persons, things, or objects under study. To study the population, we select a **sample**. The idea of **sampling** is to select a portion (or subset) of the larger population and study that portion (the sample) to gain information about the population. Data are the result of sampling from a population.

Because it takes a lot of time and money to examine an entire population, sampling is a very practical technique. If you wished to compute the overall grade point average at your school, it would make sense to select a sample of students who attend the school. The data collected from the sample would be the students' grade point averages. In presidential elections, opinion poll samples of 1,000–2,000 people are taken. The opinion poll is supposed to represent the views of the people in the entire country. Manufacturers of canned carbonated drinks take samples to determine if a 16 ounce can contains 16 ounces of carbonated drink.
From the sample data, we can calculate a statistic. A **statistic** is a number that represents a property of the sample. For example, if we consider one math class to be a sample of the population of all math classes, then the average number of points earned by students in that one math class at the end of the term is an example of a statistic. The statistic is an estimate of a population parameter. A **parameter** is a numerical characteristic of the whole population that can be estimated by a statistic. Since we considered all math classes to be the population, then the average number of points earned per student over all the math classes is an example of a parameter.

One of the main concerns in the field of statistics is how accurately a statistic estimates a parameter. The accuracy really depends on how well the sample represents the population. The sample must contain the characteristics of the population in order to be a **representative sample**. We are interested in both the sample statistic and the population parameter in inferential statistics. In a later chapter, we will use the sample statistic to test the validity of the established population parameter.

A **variable**, usually notated by capital letters such as X and Y, is a characteristic or measurement that can be determined for each member of a population. Variables may be **numerical** or **categorical**. **Numerical variables** take on values with equal units such as weight in pounds and time in hours. **Categorical variables** place the person or thing into a category. If we let X equal the number of points earned by one math student at the end of a term, then X is a numerical variable. If we let Y be a person's party affiliation, then some examples of Y include Republican, Democrat, and Independent. Y is a categorical variable. We could do some math with values of X (calculate the average number of points earned, for example), but it makes no sense to do math with values of Y (calculating an average party affiliation makes no sense).
Data are the actual values of the variable. They may be numbers or they may be words. **Datum** is a single value.

Two words that come up often in statistics are **mean** and **proportion**. If you were to take three exams in your math classes and obtain scores of 86, 75, and 92, you would calculate your mean score by adding the three exam scores and dividing by three (your mean score would be 84.3 to one decimal place). If, in your math class, there are 40 students and 22 are men and 18 are women, then the proportion of men students is 22402240 and the proportion of women students is 18401840. Mean and proportion are discussed in more detail in later chapters.

Task 9

Further practice: Exercises

1. Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money first year college students spend at ABC College on school supplies that do not include books. We randomly surveyed 100 first year students at the college. Three of those students spent $150, $200, and $225, respectively.

Key terms:

population   sample   parameter   statistics   variable   data Key

2. Determine what the key terms refer to in the following study. We want to know the average

The **population** is all first year students attending ABC College this term.

The **sample** could be all students enrolled in one section of a beginning statistics course at ABC College (although this sample may not represent the entire population).

The **parameter** is the average (mean) amount of money spent (excluding books) by first year college students at ABC College this term.

The **statistic** is the average (mean) amount of money spent (excluding books) by first year college students in the sample.

The **variable** could be the amount of money spent (excluding books) by one first year student. Let X = the amount of money spent (excluding books) by one first year student attending ABC College.

The **data** are the dollar amounts spent by the first year students. Examples of the data are $150, $200, and $225.

(mean) amount of money spent on school uniforms each year by families with children at Knoll Academy. We randomly survey 100 families with children in the school. Three of the families spent $65, $75, and $95, respectively.

Key

The **population** is all families with children attending Knoll Academy.

The **sample** is a random selection of 100 families with children attending Knoll Academy.

The **parameter** is the average (mean) amount of money spent on school uniforms by families with children at Knoll Academy.

The **statistic** is the average (mean) amount of money spent on school uniforms by families in the sample.

The **variable** is the amount of money spent by one family. Let X = the amount of money spent on school uniforms by one family with children attending Knoll Academy.

The **data** are the dollar amounts spent by the families. Examples of the data are $65, $75, and $95.

3. Determine what the key terms refer to in the following study.

A study was conducted at a local college to analyze the average cumulative GPA's of students who graduated last year. Fill in the letter of the phrase that best describes each of the items below.

1. Population _____   2. Statistic _____   3. Parameter _____ 4. Sample _____
5. Variable _____   6. Data _____

a) all students who attended the college last year
b) the cumulative GPA of one student who graduated from the college last year
c) 3.65, 2.80, 1.50, 3.90
d) a group of students who graduated from the college last year, randomly selected
e) the average cumulative GPA of students who graduated from the college last year
f) all students who graduated from the college last year
g) the average cumulative GPA of students in the study who graduated from the college last year

Key

a) all students who attended the college last year **DOES NOT APPLY**
b) the cumulative GPA of one student who graduated from the college last year **VARIABLE**
c) 3.65, 2.80, 1.50, 3.90 **DATA**
d) a group of students who graduated from the college last year, randomly selected **SAMPLE**
e) the average cumulative GPA of students who graduated from the college last year **PARAMETER**
f) all students who graduated from the college last year **POPULATION**
g) the average cumulative GPA of students in the study who graduated from the college last year **STATISTICS**

4. Determine what the key terms refer to in the following study.

As part of a study designed to test the safety of automobiles, the National Transportation Safety Board collected and reviewed data about the effects of an automobile crash on test dummies. Here is the criterion they used:

Speed at which Cars Crashed   Location of "drive" (i.e. dummies)

35 miles/hour   Front Seat

Cars with dummies in the front seats were crashed into a wall at a speed of 35 miles per hour. We want to know the proportion of dummies in the driver's seat that would have had head injuries, if they had been actual drivers. We start with a simple random sample of 75 cars.

Answers:

> The **population** is all cars containing dummies in the front seat.
>
> The **sample** is the 75 cars, selected by a simple random sample.
>
> The **parameter** is the proportion of driver dummies (if they had been real people) who would have suffered head injuries in the population.
>
> The **statistic** is proportion of driver dummies (if they had been real people) who would have suffered head injuries in the sample.
>
> The **variable** X = the number of driver dummies (if they had been real people) who would have suffered head injuries.
>
> The **data** are either: yes, had head injury, or no, did not.

5. Determine what the key terms refer to in the following study.

An insurance company would like to determine the proportion of all medical doctors who have been involved in one or more malpractice lawsuits. The company selects 500 doctors at random from a professional directory and determines the number in the sample who have been involved in a malpractice lawsuit.

Answers:

> The **population** is all medical doctors listed in the professional directory.
>
> The **parameter** is the proportion of medical doctors who have been involved in one or more malpractice suits in the population.
>
> The **sample** is the 500 doctors selected at random from the professional directory.
>
> The **statistic** is the proportion of medical doctors who have been involved in one or more malpractice suits in the sample.
>
> The **variable** X = the number of medical doctors who have been involved in one or more malpractice suits.
>
> The **data** are either: yes, was involved in one or more malpractice lawsuits, or no, was not.

6. Use the following information to answer the next five exercises.

Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average (mean) length of time in months patients live once they start the treatment. Two researchers each follow a different set of 40 patients with AIDS from the start of treatment until their deaths. The following data (in months) are collected.

Researcher A:
3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34

Researcher B:
3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

Determine what the key terms refer to in the example for Researcher A.
- population: AIDS PATIENTS
- sample: 40 PATIENTS WITH AIDS FROM THE START OF TREATMENT UNTIL THEIR DEATHS
- parameter: The average length of time (in months) AIDS patients live after treatment.
- statistic: The average length of time (in months) AIDS patients in the study live after treatment.
- variable: X= THE LENGTH OF TIME (IN MONTHS) AIDS PATIENTS LIVE AFTER TREATMENT

7. For each of the following exercises, identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.

a. Ski resorts are interested in the mean age that children take their first ski and snowboard lessons. They need this information to plan their ski classes optimally.

a. all children who take ski or snowboard lessons
b. a group of these children
c. the population mean age of children who take their first snowboard lesson
d. the sample mean age of children who take their first snowboard lesson
e. X = the age of one child who takes his or her first ski or snowboard lesson
f. values for X, such as 3, 7, and so on

b. Insurance companies are interested in the mean health costs each year of their clients, so that they can determine the costs of health insurance.

a. the clients of the insurance companies
b. a group of the clients
c. the mean health costs of the clients
d. the mean health costs of the sample
e. X = the health costs of one client
f. values for X, such as 34, 9, 82, and so on

c. A marriage counselor is interested in the proportion of clients she counsels who stay married.

a. all the clients of this counselor
b. a group of clients of this marriage counselor
c. the proportion of all her clients who stay married
d. the proportion of the sample of the counselor's clients who stay married
e. X = the number of couples who stay married
f. yes, no

d. A marketing company is interested in the proportion of people who will buy a particular product.

a. all people (maybe in a certain geographic area, such as the United States)
b. a group of the people
c. the proportion of all people who will buy the product
d. the proportion of the sample who will buy the product
e. X = the number of people who will buy it
f. buy, not buy

8. Use the following information to answer the next three exercises: A Lake Tahoe Community College instructor is interested in the mean number of days Lake Tahoe Community College math students are absent from class during a quarter.

8.1 What is the population she is interested in?

1. all Lake Tahoe Community College students
2. all Lake Tahoe Community College English students
3. all Lake Tahoe Community College students in her classes
4. all Lake Tahoe Community College math students

8.2 Consider the following:

XX = number of days a Lake Tahoe Community College math student is absent
In this case, X is an example of a:

1. variable.
2. population.
3. statistic.
4. data.

8.3 The instructor's sample produces a mean number of days absent of 3.5 days. This value is an example of a:

1. parameter.
2. data.
3. statistic.
4. variable.

a. all students who attended the college last year DOES NOT APPLY
b. the cumulative GPA of one student who graduated from the college last year VARIABLE
c. 3.65, 2.80, 1.50, 3.90 DATA
d. a group of students who graduated from the college last year, randomly selected SAMPLE
e. the average cumulative GPA of students who graduated from the college last year PARAMETER
f. all students who graduated from the college last year POPULATION
g. the average cumulative GPA of students in the study who graduated from the college last year STATISTICS

Further practice

9. Determine what the key terms refer to in the following study.

As part of a study designed to test the safety of automobiles, the National Transportation Safety Board collected and reviewed data about the effects of an automobile crash on test dummies. Here is the criterion they used:

Speed at which Cars Crashed	**Location of "drive" (i.e. dummies)**
35 miles/hour	**Front Seat**

Cars with dummies in the front seats were crashed into a wall at a speed of 35 miles per hour. We want to know the proportion of dummies in the driver's seat that would have had head injuries, if they had been actual drivers. We start with a simple random sample of 75 cars.

Answers:

The **population** is all cars containing dummies in the front seat.

The **sample** is the 75 cars, selected by a simple random sample.

The **parameter** is the proportion of driver dummies (if they had been real people) who would have suffered head injuries in the population.

The **statistic** is proportion of driver dummies (if they had been real people) who would have suffered head injuries in the sample.

The **variable** X = the number of driver dummies (if they had been real people) who would have suffered head injuries.

The **data** are either: yes, had head injury, or no, did not.

10. Determine what the key terms refer to in the following study.

An insurance company would like to determine the proportion of all medical doctors who have been involved in one or more malpractice lawsuits. The company selects 500 doctors at random from a professional directory and determines the number in the sample who have been involved in a malpractice lawsuit.

Answers:

The **population** is all medical doctors listed in the professional directory.

The **parameter** is the proportion of medical doctors who have been involved in one or more malpractice suits in the population.

The **sample** is the 500 doctors selected at random from the professional directory.

The **statistic** is the proportion of medical doctors who have been involved in one or more malpractice suits in the sample.

The **variable** X = the number of medical doctors who have been involved in one or more malpractice suits.

The **data** are either: yes, was involved in one or more malpractice lawsuits, or no, was not.

11. Use the following information to answer the next five exercises.

Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average (mean) length of time in months patients live once they start the treatment. Two researchers each follow a different set of 40 patients with AIDS from the start of treatment until their deaths. The following data (in months) are collected.

Researcher A:
3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34

Researcher B:
3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

Determine what the key terms refer to in the example for Researcher A.

- population: AIDS PATIENTS
- sample: 40 PATIENTS WITH AIDS FROM THE START OF TREATMENT UNTIL THEIR DEATHS
- parameter: The average length of time (in months) AIDS patients live after treatment.
- statistic: The average length of time (in months) AIDS patients in the study live after treatment.
- variable: X= THE LENGTH OF TIME (IN MONTHS) AIDS PATIENTS LIVE AFTER TREATMENT

12. For each of the following exercises, identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.

12.1 Ski resorts are interested in the mean age that children take their first ski and snowboard lessons. They need this information to plan their ski classes optimally.

ANSWERS:

a. all children who take ski or snowboard lessons
b. a group of these children
c. the population mean age of children who take their first snowboard lesson
d. the sample mean age of children who take their first snowboard lesson
e. X = the age of one child who takes his or her first ski or snowboard lesson
f. values for X, such as 3, 7, and so on

12.2 Insurance companies are interested in the mean health costs each year of their clients, so that they can determine the costs of health insurance.

ANSWERS:

> a. the clients of the insurance companies
> b. a group of the clients
> c. the mean health costs of the clients
> d. the mean health costs of the sample
> e. X = the health costs of one client
> f. values for X, such as 34, 9, 82, and so on

12.3 A marriage counselor is interested in the proportion of clients she counsels who stay married.

ANSWERS:

> a. all the clients of this counselor
> b. a group of clients of this marriage counselor
> c. the proportion of all her clients who stay married
> d. the proportion of the sample of the counselor's clients who stay married
> e. X = the number of couples who stay married
> f. yes, no

12.4 A marketing company is interested in the proportion of people who will buy a particular product.

ANSWERS:

> a. all people (maybe in a certain geographic area, such as the United States)
> b. a group of the people
> c. the proportion of all people who will buy the product
> d. the proportion of the sample who will buy the product
> e. X = the number of people who will buy it
> f. buy, not buy

13. Use the following information to answer the next three exercises: A Lake Tahoe Community College instructor is interested in the mean number of days Lake Tahoe Community College math students are absent from class during a quarter.

13.1 What is the population she is interested in?

1. all Lake Tahoe Community College students
2. all Lake Tahoe Community College English students
3. all Lake Tahoe Community College students in her classes
4. all Lake Tahoe Community College math students

13.2 Consider the following:

XX = number of days a Lake Tahoe Community College math student is absent
In this case, X is an example of a:

1. variable.
2. population.
3. statistic.
4. data.

13.3 The instructor's sample produces a mean number of days absent of 3.5 days. This value is an example of a:

1. parameter.
2. data.
3. statistic.
4. variable.

Measures of Central Tendency

It is also useful to be able to describe the characteristics of a distribution more precisely. Here we look at how to do this in terms of two important characteristics: their central tendency and their variability.

Central Tendency

The central tendency of a distribution is its middle—the point around which the scores in the distribution tend to cluster. (Another term for central tendency is average.) Looking back at Figure 12.1, for example, we can see that the self-esteem scores tend to cluster around the values of 20 to 22. Here we will consider the three most common measures of central tendency: the mean, the median, and the mode.

The mean of a distribution (symbolized M) is the sum of the scores divided by the number of scores. As a formula, it looks like this:

$$M = \Sigma X / N$$

In this formula, the symbol Σ (the Greek letter sigma) is the summation sign and means to sum across the values of the variable X. N represents the number of scores. The mean is by far the most common measure of central tendency, and there are some good reasons for this. It usually provides a good indication of the central tendency of a distribution, and it is easily understood by most people. In addition, the mean has statistical properties that make it especially useful in doing inferential statistics.

An alternative to the mean is the median. The median is the middle score in the sense that half the scores in the distribution are less than it and half are greater than it. The simplest way to find the median is to organize the scores from lowest to highest and locate the score in the middle. Consider, for example, the following set of seven scores:

8 4 12 14 3 2 3

To find the median, simply rearrange the scores from lowest to highest and locate the one in the middle.

2 3 3 4 8 12 14

In this case, the median is 4 because there are three scores lower than 4 and three scores higher than 4. When there is an even number of scores, there are two scores in the middle of the distribution, in which case the median is the value halfway between them. For example, if we were to add a score of 15 to the preceding data set, there would be two scores (both 4 and 8) in the middle of the distribution, and the median would be halfway between them (6).

One final measure of central tendency is the mode. The mode is the most frequent score in a distribution. In the self-esteem distribution presented in Table 12.1 and Figure 12.1, for example, the mode is 22. More students had that score than any other. The mode is the only measure of central tendency that can also be used for categorical variables.

In a distribution that is both unimodal and symmetrical, the mean, median, and mode will be very close to each other at the peak of the distribution. In a bimodal or asymmetrical distribution, the mean, median, and mode can be quite different. In a bimodal distribution, the mean and median will tend to be between the peaks, while the mode will be at the tallest peak. In a skewed distribution, the mean will differ from the median in the direction of the skew (i.e., the direction of the longer tail). For highly skewed distributions, the mean can be pulled so far in the direction of the skew that it is no longer a good measure of the central tendency

of that distribution. Imagine, for example, a set of four simple reaction times of 200, 250, 280, and 250 milliseconds (ms). The mean is 245 ms. But the addition of one more score of 5,000 ms—perhaps because the participant was not paying attention—would raise the mean to 1,445 ms. Not only is this measure of central tendency greater than 80% of the scores in the distribution, but it also does not seem to represent the behaviour of anyone in the distribution very well. This is why researchers often prefer the median for highly skewed distributions (such as distributions of reaction times).

Keep in mind, though, that you are not required to choose a single measure of central tendency in analyzing your data. Each one provides slightly different information, and all of them can be useful.

Information display

One way to display the distribution of a variable is in a frequency table. Below, for example, is a frequency table (Table 1) showing a hypothetical distribution of scores on the Rosenberg Self-Esteem Scale for a sample of 40 college students. The first column lists the values of the variable—the possible scores on the Rosenberg scale—and the second column lists the frequency of each score. This table shows that there were three students who had self-esteem scores of 24, five who had self-esteem scores of 23, and so on. From a frequency table like this, one can quickly see several important aspects of a distribution, including the range of scores (from 15 to 24), the most and least common scores (22 and 17, respectively), and any extreme scores that stand out from the rest. There are a few other points worth noting about frequency tables. First, the levels listed in the first column usually go from the highest at the top to the lowest at the bottom, and they usually do not extend beyond the highest and lowest scores in

the data. For example, although scores on the Rosenberg scale can vary from a high of 30 to a low of 0, Table 1 only includes levels from 24 to 15 because that range includes all the scores in this particular data set. Second, when there are many different scores across a wide range of values, it is often better to create a grouped frequency table, in which the first column lists ranges of values and the second column lists the frequency of scores in each range. Table 12.2, for example, is a grouped frequency table showing a hypothetical distribution of simple reaction times for a sample of 20 participants. In a grouped frequency table, the ranges must all be of equal width, and there are usually between five and 15 of them. Finally, frequency tables can also be used for categorical variables, in which case the levels are category labels. The order of the category labels is somewhat arbitrary, but they are often listed from the most frequent at the top to the least frequent at the bottom.

Reaction time (ms)   Frequency

Reaction time (ms)	Frequency
241–260	1
221–240	2
201–220	2
181–200	9
161–180	4
141–160	2

Histograms

A histogram is a graphical display of a distribution. It presents the same information as a frequency table but in a way that is even quicker and easier to grasp. The histogram in Figure 6 presents the distribution of self-esteem scores in Table 1 above.. The x-axis of the histogram represents the variable and the y-axis represents frequency. Above each level of the variable on the x-axis is a vertical bar that represents the number of individuals with that score. When the variable is quantitative, as in this example, there is usually no gap between the bars. When the variable is categorical, however, there is usually a small gap between them. (The gap at 17 in this histogram reflects the fact that there were no scores of 17 in this data set.)
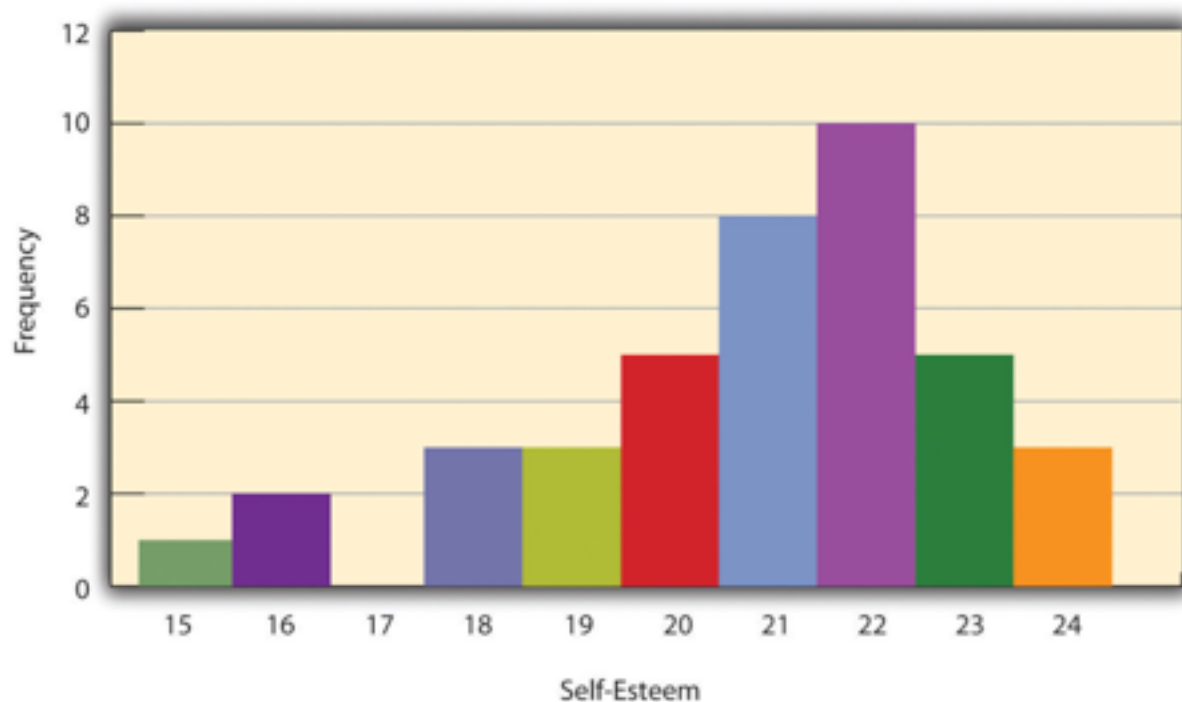

Figure 6: Histogram Showing the Distribution of Self-Esteem Scores Presented in Table 1

Distribution Shapes

When the distribution of a quantitative variable is displayed in a histogram, it has a shape. The shape of the distribution of self-esteem scores in Figure 6 is typical. There is a peak somewhere near the middle of the distribution and "tails" that taper in either direction from the peak. The distribution of Figure 6 is unimodal, meaning it has one distinct peak, but distributions can also be bimodal, meaning they have two distinct peaks. Figure 7, for example, shows a hypothetical bimodal distribution of scores on the Beck Depression Inventory.
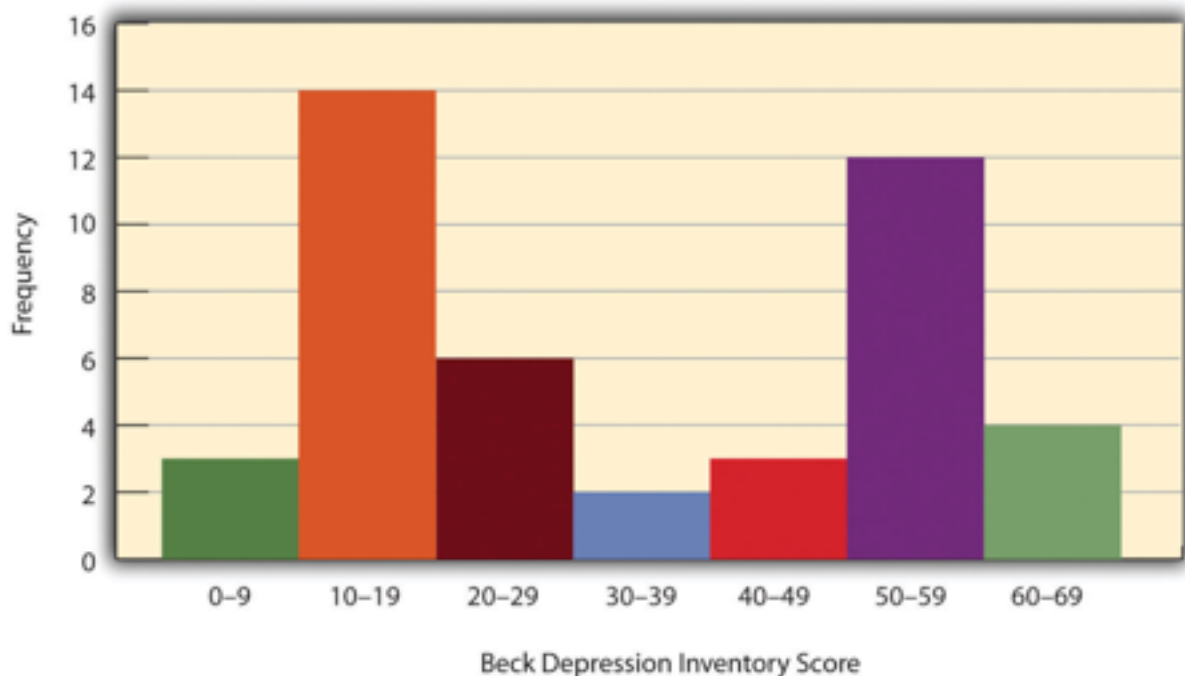

Figure 7. Histogram Showing a Hypothetical Bimodal Distribution of Scores on the Beck Depression Inventory

Another characteristic of the shape of a distribution is whether it is symmetrical or skewed. The distribution in the centre of Figure 8 is symmetrical. Its left and right halves are mirror images of each other. The distribution on the left is negatively skewed, with its peak shifted toward the upper end of its range and a relatively long negative tail. The distribution on the right is positively skewed, with its peak toward the lower end of its range and a relatively long positive tail.

Figure 8. Histograms Showing Negatively Skewed, Symmetrical, and Positively Skewed Distributions

After you administer an assessment, basic statistics give you an idea of how your students are performing and information about the test in general. Two useful statistics to measure are the mean score and the pass/fail rate. The mean is the mathematical average of all the test takers' scores; you can compute the mean by adding all the test takers' scores and dividing the total by the number of students who took the test. You can compute a mean score for the whole test and for each section. The mean score provides information about the reliability of a test over time, especially if the delivery of the course remains the same. If a test is reliable, you would expect a similar result among students at different times or between groups of similar students.

If you have a test in which various skills are tested, such as the one in the example of the test specifications with a listening, reading, and writing section, computing the mean for each section would give you an idea of which skill areas the students have more difficulty with, as indicated by a lower mean score. To determine the pass/fail rate, meanwhile, you must first decide what a passing score is on the exam. I have been in places where 60 percent is passing at the university level for some courses, depending on their difficulty, and other places where 75 percent or even higher represents the passing score. It depends on the level at which students are expected to master the course objectives, and on whether the purpose of testing is just to measure learning objectives or to move only the top performers on to the next course or level of study. Once you have determined a passing score, add the number of students who reached the passing score or higher and divide by the total number of test takers. This number represents the pass rate. The same can be done for the fail rate, counting instead the number of students who did not receive at least the passing mark and dividing by the total number of test takers. The pass/fail rate can be calculated for

the test as a whole, for each section or skill area, and even for individual questions. Looking at individual-item pass rates gives you an idea of the difficulty of each question and can help you find items that may need to be rewritten or areas that will need further instruction or review. If you add this information to the questions that you created on notecards or a computer spreadsheet, in the future you can also put together tests with questions that have a variety of difficulty levels. This information will also tell you which skills you might not need to spend so much time reviewing in the future and which ones may need more instruction time.

Task 10

Below are the data from 14 students and their score in a 10-item reading test designed by the course teacher.

a) Transfer the information on to an Excel spreadsheet.
b) Apply descriptive statistics of central tendency.
c) Calculate the overall score and mark for each student.
d) Illustrate the results in a frequency table and a histogram.
e) What inferences can you draw for each item and for the test as a whole?

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Student 1	2	6	3	4	2	3	4	3	5,5	7
Student 2	0	9	1	4	0,5	3	3	3	6,5	8
Student 3	5	7	3,5	4	0	1	0	0	5	12
Student 4	6	7	4	3,5	3,5	4	4	4	10,5	11
Student 5	4	6	4	3,5	4	4	2	3	11,5	12
Student 6	6	6	3	3,5	0	0	3,5	2	2	8,5
Student 7	4	7	5	4	2,5	3	4	3	12	13
Student 8	6	6	4,5	3	4	4	4	3	12	13
Student 9	4	9	4	2	3	1	1	0	6	10
Student 10	5,5	7	3	4	3,5	3	0	0	2	11
Student 11	4,5	5	4	4	2	3	2	0	7,5	6
Student 12	2	9	5	4	2	4	4	4	12	11
Student 13	6	9	5	4	4	3	4	4	12	14
Student 14	6	9	4	4	4	3,5	3	4	8,5	11

Closing: Summary of key concepts.

Average also called mean; a number that describes the central tendency of the data

Categorical Variable: variables that take on values that are names or labels

Data: a set of observations (a set of possible outcomes); most data can be put into two groups: **qualitative** (an attribute whose value is indicated by a label) or **quantitative** (an attribute whose value is indicated by a number). Quantitative data can be separated into two subgroups: **discrete** and **continuous**. Data is discrete if it is the result of counting (such as the number of students of a given ethnic group in a class or the number of books on a shelf). Data is continuous if it is the result of measuring (such as distance traveled or weight of luggage)

Numerical Variable: variables that take on values that are indicated by numbers

Parameter: a number that is used to represent a population characteristic and that generally cannot be determined easily

Population: all individuals, objects, or measurements whose properties are being studied

Probability: a number between zero and one, inclusive, that gives the likelihood that a specific event will occur

Proportion: the number of successes divided by the total number in the sample

Representative Sample: a subset of the population that has the same characteristics as the population

Sample: a subset of the population studied

Statistic: a numerical characteristic of the sample; a statistic estimates the corresponding population parameter.

Variable: a characteristic of interest for each person or object in a population

Source: **https://cnx.org/contents/MBiUQmmY@19.6:y0GFmfab@8/Definitions-of-Statistics-Prob**
This work is licensed by OpenStax under a **Creative Commons Attribution License (by 4.0).**

UNIT 3:
Assessing productive and interactive skills

Learning Outcome

As in Unit 2, you are expected to plan, design, apply and reflect on two assessment procedures. This time, however, the focus will be on speaking and writing. To help you in the process, you will read about alignments, oral and written assessment, and classroom assessment techniques. You are expected to design assessment procedures and instruments. You are also expected to assess spoken performance by using rubrics, demonstrate inter-rater reliability and give pertinent feedback to students.

Evidence of learning: Student creates validated rubrics for writing assignments and uses rubrics for spoken performance and provide qualitative feedback in tutorial sessions.

Before we start!

Task 11

a) Search on the internet samples of (in an ELT context):
- Checklists
- Rating scales
- Task-specific scales
- Holistic scales
- Analytic scales

b) Be ready to discuss with the class what purpose they serve (formative and/or achievement assessment), advantages and disadvantages, elements that are being assessed both for writing and speaking performance (task description, criteria, levels and scoring scale, standards for performance / descriptors).

Assessing Speaking: An overview.

Green (2014) refers to Cutler's (2005) observation on productive and interactive language skills. Green observed that in some ways, productive language processing is similar to receptive processing -listening and reading. In the former, the speaker or writer begins with a message to be conveyed and the process of production implies putting the message into speech or writing. By contrast, the listener or reader begins with uncertainty about the message. and must 'figure out' the visual or auditory signal to interpret the message. However, either as speaker/writer or listener/reader, people have goals in their minds, in other words, a message they want / need to convey. To achieve that, they draw on their mental store of vocabulary and on knowledge of grammar and discourse. The addressers need to make decisions on how best to formulate the message to convey it to the intended audience. This will involve utilising knowledge of the world, knowledge of the topic, about the audience, the context in which the communication occurs and about the types of genres of speech or writing that are suited to the given context.

Brown (2004) presents a five-level taxonomy for speaking assessment:

1. Imitative: Imitative speaking sits at one end of the continuum. It is basically being able to imitate / reproduce a word or short utterance with a clear focus on prosodic features.

2. Intensive: It refers to the production of short stretches of language designed to showcase competence in a narrow band of grammatical, phrasal, lexical or phonological relationships. Semantics play a role but interactions ar minimal. Examples of intensive assessment tasks include directed response tasks, reading aloud, sentence and dialogue completion, limited picture-cued tasks.

3. Responsive: Responsive assessment tasks include interaction and test comprehension within short conversations, standard greetings and small talk, simple requests, and comments.

4. Interactive: The responses at this level of the taxonomy are are longer and more complex than at the responsive level. These include multiple exchanges and/or multiple participants. Interactions can be of two types: transactional and interpersonal. Transactional language pursues exchanging specific information whereas interpersonal language maintain social relationships. In the latter, oral production can become pragmatically more complex as it may include casual register: colloquial language, ellipsis, slang, humour and other sociolinguistic conventions.

5. Extensive. This includes speeches, oral presentations, and story-telling. A certain level of planning is involved and tends to be more formal or academic. That said, extensive speaking may also include informal monologues such as casually delivered speech (telling an anecdote, recounting the plot of a movie or novel).

What speaking entails

Organizing

Source: **http://www.learnalberta.ca/content/eslapb/strands_speaking.html**

Speaking Competency is categorized into four categories; linguistic competence, strategic competence, socio-linguistic competence, discourse competence. These categories assist teachers to organize the lesson plans they create to ensure the success of the ESL students.

Effective lesson plans for ESL students are done in both individual situations and group situations. Formative assessment is key in order to gage where on the ESL benchmarks the student should be placed.

For Individual formative assessment

- Be sure to include one-on-one time with the student during class time
- Use video and/or audio recording to help the student gain confidence
- During verbal reading assessment ensure there is effective interaction
- Conduct it on an ongoing bases
- Use prior knowledge and relevant information from the students' backgrounds
- Use visual cues

For Group formative Assessment

- Teachers should closely observe the student during class group work
- Encourage the student to participate within discussions
- Place the student in groups that involve other students who are supportive of their needs

How to conduct assessment:

- Interviews
- Oral Reports
- Role Plays
- Describing/Explaining
- Paraphrasing stories

Task 12

Hands-on activity.

- Arrange to attend as a co-assessor to one of the oral assessment procedures either at your school or at University.

- Conference with the class teacher to find out about GTI -General Test Information- and logistics for attending as a co-assessor.

- Help out with GTI, materials and classroom layout for the speaking assessment.

- During assessment, take notes on students' performance, seeing to the different criteria specified in the rubric.

- Participate in the qualitative feedback tutorials at the end of each presentation.

- Conference with the other assessors on setting the standard and granting a mark to each of the participants.

- Write a reflective-log entry on this learning experience.

Assessing writing: an overview.

Obviously making the switch from another language to English is going to be an intimidating task for the student as well as the teacher. It is important for students to be able to read, understand and speak in English and it is also essential for the student to be able to express themselves through writing as well. Assessment of ESL writing is a task that may seem daunting to some teachers, but it is important to know that proper assessment is a process that takes time, patience and the proper use of steps and scaffolding to achieve this goal.

Types of writing performance

Brown (2004) presents four categories of written performance that encompass a range of written production. Each category is similar to the categories presented and defined for reading, listening and speaking but, as for the others, these categories reflect the uniqueness of the skill area.

1. Imitative:This category includes the ability to spell correctly, demonstration of phoneme-grapheme correspondences in the English spelling system, mechanics of writing. Form is the primary focus while meaning and context play a secondary role. Some examples of imitative exercises and tasks are: listening cloze, picture-cued tasks, form completion tasks, converting numbers and abbreviations to words.

2. Intensive (controlled): Meaning and context are of some importance in determining correctness and appropriateness but most assessment tasks are still focused on form at sentence level. Some examples of intensive exercises and tasks are: dictation and dicto-comp, grammatical transformation tasks, picture-cued tasks (short sentences, picture description, picture sequence description) short answers, sentence completion tasks.

3. Responsive: Learners are expected to perform at a limited discourse level, connecting sentences into a sequence of 2 to 3 paragraphs. Some genres include brief narratives and descriptions, short reports, summaries, brief responses to reading. At this level, learners are expected to have mastered the fundamentals of sentence-level grammar and hence, is more focused on discourse conventions. Form-focused attention is mostly at the discourse level: meaning and context play a key role, much more than in the previous category. Some examples of intensive exercises and tasks are:paraphrasing, guided question and answer, paragraph construction tasks).

4. Extensive: Learners at this level must show successful management of processes and strategies of writing for a wide range of purposes. Writers are able to focus on achieving a purpose, organising and developing ideas logically, demonstrating syntactic and lexical variety.

Areas of Assessment

- **Linguistic Competence:** It includes grammar, vocabulary and the mechanics of language. These performance aspects of written language are highly valued in judging the quality of a piece of writing.

- **Strategic Competence:** A strategic writer creates text appropriate to purpose, voice, audience, form and occasion.

- **Socio-linguistic Competence:** The student is able to vary the use of language in terms of audience, purpose, genre, topic and degree of formality. It is through socio-linguistic competence that writers maintain their credibility with their audience.

- **Discourse Competence:** Knowledge of transitional devices that connect one element in a text with another. Transitional devices such as reference, substitution, ellipsis, conjunction and lexical cohesion.

Assessment Tools

For a teacher it is important to keep the process of formative assessment in mind. By using formative assessment it allows the teacher to know where the student needs to put in some extra work and also shows where students are being successful. Focussing on the small goals and using steps to success will ensure a positive and motivating learning environment for the students.

Some interesting ideas for assessment tools are:

- Learning logs
- Dialogue journals
- Summaries
- Unedited student written work
- Close procedures
- Using samples of completed work to help improve student understanding

Some assessment ideas to keep in mind when reviewing students work are:

Does the student:

- Write for a variety of audiences and purposes?
- Write on specific topics?
- Demonstrate a variety of written formats?
- Manipulate verb tenses?
- Vary sentence construction?
- Pay attention to punctuation and spelling?

Providing Feedback

Providing feedback for writing assessment is as crucial -if not more- than designing the task and assessment instrument (rubric, rating scale, checklist).

Once basic way of providing feedback is using a writing correction code as the one in Fig.9 below.

Code	Type of mistake	Example
GR	Grammar	People is becoming more and more aware of gender pay gap.
TN	Tense	As I drove closer to the edge of town, the landscape was changing dramatically.
SP	Spelling	His lack of responsalibity is what explains his being sacked.
P	Punctuation/Capitalization	I love learning english
PREP	Preposition	I will pop up to get some coffee. Fancy one?
WW	Wrong word	I was taken aback by his utter carefulness.
WO	Wrong order	Not only you need cash but also a robust credit line.
Ù	Word missing	My dad would like to go to the stadium with him but I just can't stand the idea of sitting there looking at a game that bores me to death.
SYN	Provide a synonym for this word	The size of the rock was very, very big.
RW	Try re-writing	I just don't get the point of wrecking my nerves for to get so little.
REG	Wrong register	So, Mr Bond. Do you want a Martini?
???	Ideas are not clear	What I do want to say you cannot remind unmoved by the colossal error. It is nothing to worry too much.

For more qualitative feedback on writing, read the article entitled "The impact of oral and written feedback on EFL writers with the use of Screencast", http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-07902016000200006

Task 13

Your teacher will provide you with a task description and some first drafts from an upper-intermediate level class. You are expected to:

a) create a holistic scale for an overall assessment of the task
b) provide basic feedback using the writing correction code (Fig. 9)
c) use the holistic scale to provide quantitative and qualitative feedback.
d) Write a reflective-log entry on this learning experience.

Closing: Summary of key concepts.

Checklist

A checklist is a tool for identifying the presence or absence of conceptual knowledge, skills, or behaviours. Checklists are used for identifying whether key tasks in a procedure, process, or activity have been completed. The tasks may be a sequence of steps or include items to verify that the correct sequence was followed. You may need to observe the tasks being followed because, in general, you cannot judge what tasks the learner did from the end product. Remember that some attitudes may be indirectly observed. For example, safety attitudes can be observed by seeing if safety equipment is worn.

A checklist may also be given to students to follow in completing a procedure (e.g., in a shop or lab). A checklist itemizes task descriptions in one column and provides a space beside each item in a second column to check off the completion of the task.

Characteristics of checklists. Checklists should:

- have criteria for success based on expected outcomes
- be short enough to be practical (e.g., one sheet of paper)
- have tasks chunked into logical sections or flow from start to finish
- highlight critical tasks
- have sign-off points that prevent students from proceeding without approval, if needed
- be written with clear, detailed wording to minimize the risk of misinterpretation
- have space for other information such as the student's name, date, course, examiner, and overall result
- be reviewed by other instructors

Rating scales

What is a rating scale?

A rating scale is a tool used for assessing the performance of tasks, skill levels, procedures, processes, qualities, quantities, or end products, such as reports, drawings, and computer programs. These are judged at a defined level within a stated range. Rating scales are similar to checklists except that they indicate the degree of accomplishment rather than just yes or no.

Rating scales list performance statements in one column and the range of accomplishment in descriptive words, with or without numbers, in other columns. These other columns form "the scale" and can indicate a range of achievement, such as from poor to excellent, never to always, beginning to exemplary, or strongly disagree to strongly agree.

Some tasks, such as procedures and processes, need to be observed in order to be assessed.

Characteristics of rating scales. Rating scales should:

- have criteria for success based on expected outcomes
- have clearly defined, detailed statements

For assessing end products, it can sometimes help to have a set of photographs or real samples that show the different levels of achievement. Students can visually compare their work to the standards provided.

- have statements that are chunked into logical sections or flow sequentially
- include clear wording with numbers when a number scale is used

As an example, when the performance statement describes a behaviour or quality, 1 = poor through to 5 = excellent is better than 1 = lowest through to 5 = highest or simply 1 through 5.

The range of numbers should be the same for all rows within a section (such as all being from 1 to 5). The range of numbers should always increase or always decrease. For example, if the last number is the highest achievement in one section, the last number should be the highest achievement in the other sections.

- have specific, clearly distinguishable terms

Using good then excellent is better than good then very good because it is hard to distinguish between good and very good. Some terms, such as often or sometimes, are less clear than numbers, such as 80% of the time.

- be short enough to be practical
- highlight critical tasks or skills
- indicate levels of success required before proceeding further, if applicable
- sometimes have a column or space for providing additional feedback
- have space for other information such as the student's name, date, course, examiner, and overall result
- be reviewed by other instructors

Considerations for numeric rating scales

If you assign numbers to each column for marks, consider the following:

- What should the first number be? If 0, does the student deserve 0%? If 1, does the student deserve 20% (assuming 5 is the top mark) even if he/she has done extremely poorly?

- What should the second number be? If 2 (assuming 5 is the top mark), does the person really deserve a failing mark (40%)? This would mean that the first two or three columns represent different degrees of failure.

- Consider variations in the value of each column. Assuming 5 is the top mark, the columns could be valued at 0, 2.5, 3, 4, and 5.

- Consider the weighting for each row. For example, for rating a student's report, should the introduction, main body, and summary be proportionally rated the same? Perhaps, the main body should be valued at five times the amount of the introduction and summary. A multiplier or weight can be put in another column for calculating a total mark in the last column.

- Consider having students create the rating scale. This can get them to think deeply about the content.

Course Summary

Summative Assessments

Summative assessments are used to evaluate student learning, skill acquisition, and academic achievement at the conclusion of a defined instructional period—typically at the end of a project, unit, course, semester, program, or school year. Generally speaking, summative assessments are defined by three major criteria:

- The tests, assignments, or projects are used to determine whether students have learned what they were expected to learn. In other words, what makes an assessment "summative" is not the design of the test, assignment, or self-evaluation, per se, but the way it is used—i.e., to determine whether and to what degree students have learned the material they have been taught.

- Summative assessments are given at the conclusion of a specific instructional period, and therefore they are generally evaluative, rather than diagnostic—i.e., they are more appropriately used to determine learning progress and achievement, evaluate the effectiveness of educational programs, measure progress toward improvement goals, or make course-placement decisions, among other possible applications.

- Summative-assessment results are often recorded as scores or grades that are then factored into a student's permanent academic record, whether they end up as letter grades on a report card or test scores used in the college-admissions process. While summative assessments are typically a major component of the grading process in most districts, schools, and courses, not all assessments considered to be summative are graded.

Summative assessments are commonly contrasted with **formative assessments**, which collect detailed information that educators can use to improve instruction and student learning while it's happening. In other words, formative assessments are often said to be for learning, while summative assessments are of learning. Or as assessment expert Paul Black put it, "When the cook tastes the soup, that's formative assessment. When the customer tastes the soup, that's summative assessment." It should be noted, however, that the distinction between formative and summative is often fuzzy in practice, and educators may have divergent interpretations and opinions on the subject.

Some of the most well-known and widely discussed examples of summative assessments are the **standardized tests** administered by states and testing organizations, usually in math, reading, writing, and science. Other examples of summative assessments include:

- End-of-unit or chapter tests.
- End-of-term or semester tests.
- Standardized tests that are used to for the purposes of school accountability, college admissions (e.g., the SAT or ACT), or end-of-course evaluation (e.g., Advanced Placement or International Baccalaureate exams).
- Culminating d**emonstrations of learning** or other forms of "performance assessment," such as **portfolios** of **student work** that are collected over time and evaluated by teachers or **capstone projects** that students work on over extended periods of time and that they present and defend at the conclusion of a school year or their high school education.

While most summative assessments are given at the conclusion of an instructional period, some summative assessments can still be used diagnostically. For example, the growing availability of student data, made possible by online grading systems and databases, can give teachers access to assessment results from previous years or other courses. By reviewing this data, teachers may be able to identify students more likely to struggle academically in certain subject areas or with certain concepts. In addition, students may be allowed to take some summative tests multiple times, and teachers might use the results to help prepare students for future administrations of the test.

It should also be noted that districts and schools may use "interim" or "benchmark" tests to monitor the academic progress of students and determine whether they are on track to mastering the material that will be evaluated on end-of-course tests or standardized tests. Some educators consider interim tests to be formative, since they are often used diagnostically to inform instructional modifications, but others may consider them to be summative. There is ongoing debate in the education community about this distinction, and interim assessments may defined differently from place to place. See **formative assessment** for a more detailed discussion.

Formative Assessment

In an effort to support the development of a common, research-based understanding of formative assessment the Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS) published a definition of formative assessment in 2007: "Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and

learning to improve students' achievement of intended instructional outcomes." Central to this definition are several important ideas.

1. Formative assessment is not a test, assessment, or quiz given at the end of a learning period, but an ongoing process of collecting evidence of student learning during instruction to inform next steps in teaching and learning while there is still an opportunity to influence learning. Identifying areas of need at the end of a unit may influence subsequent instruction, but it is not the heart of formative assessment.

2. The idea of "during instruction" can mean both literally during a class period as students and teachers are engaged in a learning experience, and also more broadly, during an instructional sequence that may span several weeks. A teacher can make adjustments to the instructional plans to account for students' current understanding and to support them moving closer to the intended learning goals.

3. The process of formative assessment includes both students and teachers in the collection and consideration of evidence of learning; formative assessment is something teachers do with students. The FAST SCASS further expanded on this definition by identifying five attributes of effective formative assessment, listed below.

3.1. Learning Progressions. Learning progressions should clearly articulate the sub-goals of the ultimate learning goal.

3.2. Learning Goals and Criteria for Success. Learning goals and criteria for success should be clearly identified and communicated to students.

3.3. Descriptive Feedback. Students should be provided with evidence-based feedback that is linked to the intended instructional outcomes and criteria for success.

4. Self- and Peer Assessment. Both self- and peer assessment are important for providing students an opportunity to think meta-cognitively about their learning.

5. Collaboration. A classroom culture in which teachers and students are partners in learning should be established.

Techniques and instruments

1. Essays or projects

Asking students to write essays or create projects is a great way to provide an opportunity for them to summarize or apply what they have learned during a unit of study. In language classrooms, these two types of assignments require students to demonstrate their ability to use language structures or specific vocabulary in an authentic task.

Here is an example that incorporates arguments and the future tense. Students read articles that argue a position or state opinions and have mock debates as part of the unit. As a final assessment, they write an argumentative essay where they take a position on the use of mobile phones in the classroom. They choose a stance, for or against, and discuss three potential benefits or consequences using the future tense and the If, then sentence structure. In their essays, they also incorporate related vocabulary such as pro, con, advantage, disadvantage, benefit, and drawback. To assign this as a project with the same language requirements, ask students to make a poster or brochure supporting their stance.

2. Presentations, speeches, skits, or commercials

Assignments like presentations, speeches, skits, or commercials are creative tasks that ask students to incorporate what they have learned into some type of performance. They are a great way to offer students who are stronger at speaking or performing a chance to excel. These tasks can be less structured than an essay or project and allow students more creative freedom to demonstrate what they have learned. Additionally, students enjoy seeing their peers perform. Therefore, the audience can often be just as engaged in watching the performance as the presenters are in performing it! This type of assignment provides students with a chance to review information both as performers and audience members, something they would not experience by handing in an essay or taking a test.

For example, as a summative assessment for the topic of mobile phone use in the classroom, students can make a presentation or give a speech arguing for or against the use of mobile phones in the classroom.

The same requirements can still apply: students use the future tense, argue a position with three potential outcomes, use If, then structures, and incorporate some key vocabulary. Students can also create a commercial or skit with the same requirements.

3. Portfolios

Portfolios are collections of student work used to demonstrate the student's mastery of specific content. Portfolios can include classwork, homework, assessment tasks given by the teacher, peer assessments, or self-assessments conducted by the student. Pre- and post-tests, item descriptions, or student reflections may also be included.

Some portfolios aim to show student growth. These may include early assignments where it is evident that the student did not grasp the concepts being taught, and then later assignments where the student was able to perform successfully. The purpose of this type of

portfolio is to show that the student's understanding of content has increased over time. Including a pre- and post-test in a portfolio of this type is a great way to show that the student has improved. An additional benefit of this type of portfolio is that it shows students their own progress over time.

Other portfolios aim to showcase a student's best work related to a certain skill or learning outcome. In creating this type of portfolio, one would likely give his/her students a choice about which items to include. For instance, if students have completed a total of fifteen assignments related to a specific learning outcome or skill, ask them to choose ten items to include in the portfolio. Depending on the age of your students, you can also assist them with making choices about what to include.

Sometimes teachers ask students to provide a description of each item in a portfolio or even a reflection about how the piece of work demonstrates their learning or growth. This can be done with any type of portfolio. Here are some sample questions for students to answer:

- How did this activity require you to show your understanding of the concept or topic?

- If you struggled with this assignment, how well do you think you would complete it now? What would you do differently?

- How does this collection of work show that you improved your understanding of the topic over time?

- Why did you choose to include these specific assignments to demonstrate your performance?

4. Rubrics

Rubrics are an excellent tool to use with summative assessment tasks such as those described above. A rubric allows you as a teacher to clearly communicate what a highly successful performance on a task looks like, therefore giving your students specific criteria to guide their completion of an assignment. Let's revisit the example of writing an argumentative essay for or against the use of mobile phones in the classroom. Based on the requirements you have given your students, start creating a rubric by thinking about what the most successful essay would include. Then, define criteria for each subsequent level of performance on the task. Here is an example rubric.

	Stance / Position	Supporting Reasons	Vocabulary Use	Future Tense	If, then Sentences
4	Clearly stated in introduction, aligned with reasons, consistent throughout essay.	Three reasons are given and are well discussed and supported.	9-10 of the key terms are used in the essay.	Tense is correctly used and consistent throughout the essay.	There is at least one "If, then" sentence in each paragraph.
3	Clear and consistent, but may not be stated in introduction. Well-aligned with reasons.	Three reasons are given but discussion or support needs more development.	6-8 of the key terms are used in the essay.	There are fewer than five instances where the tense is not used or is used incorrectly.	One paragraph may not contain any sentences with "If, then" structure.
2	Position may not be clearly stated or not aligned with one reason.	Only two reasons are given and/or there is not enough discussion and support.	3-5 of the key terms are used in the essay.	There are fewer than ten instances where the tense is not used or is used incorrectly	Two paragraphs may not contain any sentences with "If, then" structure.
1	Position not stated or inconsistent. Not aligned with two or more reasons.	Only one reason or no reasons are given and the support is not sufficient.	Fewer than 3 of the key terms are used in the essay.	There are ten or more instances where the tense is not used or is used incorrectly.	More than two paragraphs may not contain any sentences with "If, then" structure.

Sharing the rubric with students before they begin their essays gives them exact criteria they must attend to in order to perform successfully on the assignment. Creating a rubric also helps you as a teacher to define exactly what student performance will look like at each level. This makes it easier to give students grades and enables you to be more consistent in evaluating their work.

4.1 Rubrics and formative assessment

Using the FAST SCASS definition of formative assessment, and the attributes of effective formative assessment, ten dimensions of formative assessment practice have been identified that could be observed during a lesson. The dimensions represent a set of integrated formative assessment practices. Focusing on just a single dimension likely would not result in a robust implementation of formative assessment. Rather an integrated approach is required. However, for the purpose of discussing practice it can be useful to separate them out and sometimes to focus on just a subset. They are listed below:

1. Learning Goals
2. Criteria for Success
3. Tasks and Activities that Elicit Evidence of Student Learning
4. Questioning Strategies that Elicit Evidence of Student Learning
5. Feedback Loops During Questioning
6. Descriptive Feedback
7. Peer Assessment
8. Self-Assessment
9. Collaboration
10. Using Evidence to Inform Instruction

The rubrics cluster into several groups as shown in the figure on this page. The first two dimensions focus on information teachers provide or develop with students about what the learning will be, or how teachers and students will know when it has been understood. They help teachers and students identify where they are headed. The next two dimensions focus on ways of collecting evidence of student learning: through tasks and activities designed to elicit evidence of student thinking, through deliberate and planned questioning strategies; and through student self-assessment. These three dimensions help students and teachers understand where students are in their learning currently.

Feedback can be used to close the gap between current learning and intended learning. There are three dimensions that address distinct aspects of feedback: Feedback Loops, Individualized Descriptive Feedback, and Peer Assessment. The Feedback Loops dimension is specific to more informal feedback that often occurs in real-time during a lesson. The Individualized Descriptive Feedback dimension is specific to more formal feedback that tends to be given to individual students on a specific piece of work, either in written form or orally (e.g., during student/teacher conferences) by the teacher. The Peer Assessment dimension includes the role of student-to-student feedback. All of these dimensions center on the use of evidence to inform instruction. This work takes place in a supportive learning context, where Collaboration is valued (teacher to teacher, teacher to student, and student to student). For each of the ten dimensions a rubric is provided and observation notes addressing particular aspects of the rubric. The rubric is organized as a table with a set of columns which, reading from left to right, describe a novice or incomplete implementation to a more expert level of implementation. Each rubric describes both the teacher role in a particular formative assessment dimension and also the student role. The rubrics describe the level of implementation of a particular aspect of practice, not the level of expertise of a teacher. There are four levels or categories of implementation for each rubric. The levels are referred to both by names and by numbers to indicate a progression of skills and abilities:

- Level 1: Beginning
- Level 2: Developing
- Level 3: Progressing
- Level 4: Extending

5. Classroom assessment techniques

The following assessment techniques can be used for effective and practical measurements of students' abilities, progress, and achievement in a variety of educational settings.

5.1 Nonverbal Responses

At the early stages of learning, before the emergence of speech, children should be instructed and assessed largely through the use of physical performance responses and pictorial products (Tannenbaum 1996). These tasks require simple directions to carry out. As an assessment technique, this type of response may help lower the level of anxiety normally associated with evaluation, as students see it as a natural extension of learning activities. At a later stage, students may perform hands-on tasks. For example, they may be asked to "produce and manipulate drawings, dioramas, models, graphs, and charts" (Tannenbaum 1996:1). This technique fits very well within the Total Physical Response methodology for early language development (Asher 1988).

5.2 Oral interviews

Pierce and O'Malley (1992) suggest using visual cues in oral interviews at the early stages of acquisition. Thus a student may be asked to choose pictures to talk about, and the teacher's role is to guide the student by asking questions that require the use of related vocabulary. This technique works well during the early speech and speech emergence stages.

5.3 Role-Play

This informal assessment technique combines oral performance and physical activity. Children of all ages, when assessed through this technique, feel comfortable and motivated, especially when the activity lends itself to cooperative learning and is seen as a fun way of learning. Kelner (1993) believes that role-play can be an enjoyable way of informal assessment that could be used effectively within a content-based curriculum. For example, he recommends the use of role play to express mathematical concepts such as fractions, to demonstrate basic concepts in science such as the life cycle, and to represent historical events or literary characters.

5.4 Written Narratives

Assessment of the written communicative abilities of children could be achieved through purposeful, authentic tasks, such as writing letters to friends, writing letters to favorite television program characters, and writing and responding to invitations. Young learners enjoy storytelling and are usually motivated to listen to stories as well as to tell them. Teachers can take advantage of this interest in stories and have their students write narratives that relate to personal experiences, retell or modify nursery stories and fairy tales, or retell historical events from different perspectives. Oller (1987) suggests the use of a narrative development technique in an integrated process of teaching and assessment. The first step in the process is to check on how well learners are following the story line. To establish the basic facts, the teachers asks yes-no questions, then the teacher moves on to information questions.

5.5 Presentations

Presentations are important for assessment because they can provide a comprehensive record of students' abilities in both oral and written performance. Furthermore, presentations give the teacher some insights into student's interests, work habits, and organizational abilities. Presentations cover a wide range of meaningful activities, including poetry readings, plays, role-plays, dramatizations, and interviews. Classroom presentations are nowadays becoming more sophisticated as a result of increasing access to educational technology. In many parts of the world, students are becoming more aware of the power of multimedia for communicating information, and they enjoy keeping audio, video, and electronic records of their involvement in class presentations.

5.6 Student-teacher conferences

Student-teacher conferences, including structured interviews, can be an effective informal way of assessing a student's progress in language learning. Conferences and interviews provide opportunities for one-on-one interactions where the teacher can learn about a student's communicative abilities, emotional and social well-being, attention span, attitudes, pace of learning, and strengths and weaknesses (Smith 1996; Allerson and Grabe 1986). Conferences can be most effective when they follow focused observations. Observations could be done in class, for example, in cooperative learning groups, or out of class, for example, on the playground. Gomez, Parker, Lara-Alecio, Ochoa, and Gomez, Jr. (1996) have developed an observational instrument for assessing learners' oral performance in naturalistic language settings, which focuses on these seven language abilities: understanding by others, providing information needed by the listener, absence of hesitations, willingness

to participate in conversations, self-initiated utterances, accuracy (in grammar, usage, and vocabulary), and topic development. Tambini (1999) also recommends the use of conferences to assess the oral and written abilities of children. He, too, favors conferences that follow observations and concentrate directly on the learning processes and strategies employed by the student. For assessing oral skills, he suggests that children be evaluated primarily on their ability to understand and communicate with teachers and classmates. In assessment of writing tasks, conferences could be used to discuss drafts of essays and evaluate progress.

5.7 Learning logs

A learning log is a record of the students' experiences with the use of the English language outside the classroom, including the when and the where of language use and why certain experiences were successful and others weren't. Students may also use logs to comment on what they have studied in class and to record what they have understood and what they haven't (Brown 1998). An advantage of learning logs is that they can contribute to the teacher's understanding of the students' use of metacognitive learning strategies.

5.8 Dialogue journals

These journals are interactive in nature; they take the form of an ongoing written dialogue between teacher and student. Dialogue journals have proven effective and enjoyable for students regardless of their level of proficiency. They are informal and provide a means of free, uncensored expression, enabling students to write without worrying about being corrected (Peyton and Reed 1990). Teachers can also use journals "to collect information on students' views, beliefs, attitudes, and motivation related to a class or program or to the process involved in learning various language skills" (Brown

1998:4). As an assessment technique, dialogue journals can help the teacher assess students' writing ability and improvement over time. Peer and Group Assessment: Recent trends in EFL/ESL teaching methodology have stressed the need to develop students' ability to work cooperatively with others in groups. For assessment, for example, students can write evaluative, encouraging notes for each member of their team emphasizing their positive contribution to team work. The role of the teacher would be to provide guidance, to explain to the students what they have to evaluate in one another's work, and to help them identify and apply properly the evaluation criteria. At the end of group tasks, if necessary, the teacher can give each student a test to check their individual performance. Proponents of cooperative learning suggest the teacher should give a group grade to help reinforce the merits of group work.

Self-Assessment

Learners may also participate in self-assessment. Although self-assessment may seem inappropriate at first, it can yield accurate judgments of students' linguistic abilities, weaknesses and strengths, and improvement (McNamara and Deane 1995). Self-assessment could be done using one of the following two techniques: K-W-L charts: With this type of chart, individual students provide examples of what they know, what they wonder, what they have learned. K-W-L charts are especially effective when used at the beginning and at the end of a period of study. At the start of a course, the completed charts can help the teacher learn about students' background knowledge and interests. At the end of a course, the charts can help the students reflect on what they have learned as well as gain awareness of their improvements (Tannenbaum 1996).

https://americanenglish.state.gov/files/ae/resource_files/meaningful_assessments_week_3.pdf

Checklists, rubrics, and questionnaires offer options for learners to assess both their learning processes and language performance. Self-reflection on the learning process, or metacognitive skills, gives learners a chance to identify their own strengths and weaknesses as they learn. For example, a simple list of questions about participation and attention, where students rate their abilities on a scale of 1 to 5, can help a teacher identify what students are doing and thinking while working through the learning process.

Table 1: Simple list of questions

	1	2	3	4	5
I listen when the teacher gives directions.					
I ask questions when I don't understand.					
I participate in groups.					
I participate in class activities.					

This kind of self-reflection asks the learner to stop and think about their performance in class, and targets a specific aspect of their learning. The teacher can use this feedback to identify any discrepancies between actual student performance and student self-perceptions. Once discrepancies or similarities are recognized, teachers can then take steps to reconcile the differences and reinforce the similarities. Here's an example based on the above self-reflection. A student rates his participation high, but the teacher notices that the student rarely talks in groups or class. As a result, the teacher can see that there is a discrepancy between what the student believes he does and what he actually does. Perhaps the student

doesn't understand the expectations associated with participation or the teacher has not clearly explained these expectations. Either way, the teacher can use this information as a way to better understand the needs of the student and further support the student's success in the classroom. Perhaps the teacher can use the information to meet with the student to discuss participation expectations or spend time in class reviewing the expectations for all students. By completing regular self-reflections on process and metacognitive skills, learners feel invested in the process and also provide valuable feedback that supports teacher planning and involvement.

Another type of self-assessment gives learners a chance to assess their own language acquisition. Depending on the skill and task, a teacher can use lesson/unit objectives to create a checklist for learners to use as they review their own work, or the teacher and students can work together to develop a checklist. Here is an example of such a checklist, one created for an early draft of a paragraph writing assignment.

Table 2: Paragraph Checklist

	YES	NO
My paragraph has a topic sentence.		
My topic sentence contains a topic.		
My topic sentence contains a controlling idea.		
My paragraph has a sentence that explains how my example relates to my topic.		
My paragraph has one example of my topic and controlling idea.		
My paragraph has a concluding sentence that restates my topic sentence.		

Once students have assessed their own writing, they can be tasked with further activities related to their answers. For example, for every no answered on the checklist, the teacher asks the student to redo that part of the paragraph as a part of the revision process. This checklist can also be used again with slightly different wording for future revisions of the paragraph and can be adapted for a number of different skills and activities. For example, a speaking task might have learners record a message to set up a doctor's appointment. In the task, learners must:

- state their names
- say the reason they are calling
- ask for someone to return their call at the number given
- say thank you

After students record themselves, they can use a checklist while they listen to their own recording and see if they fulfilled the requirements of the assignment. After each student has checked his or her own recording, he or she can redo it in an effort to fix the things missed in the first recording. This type of self-assessment puts the task, learning, and revision in the hands of learners, furthering their sense of autonomy. Self-assessments of objective tasks are not limited to speaking, but can also be used for spelling, vocabulary, or grammar practice. In these types of self-assessments, learners are given the correct answers or rules and are asked to fix their own mistakes rather than relying on the teacher to make the corrections. When learners are involved in the correction process, their memory and acquisition of the skill are better enhanced and teachers are freed up to focus on other tasks.

The overarching idea in self-assessment and reflection is to involve the learner in the assessment process as a way to increase motivation and learner autonomy. As teachers, facilitating language learning is only part of our job. We also want to develop self-motivated, autonomous learners who can succeed beyond our classrooms. In assessment, we can utilize the knowledge students have about their own processes to inform and foster motivation and autonomy.

References and Acknowledgments

Brown, H. Douglas (2004). Language Assessment: principles and classroom practices. White Plains, NY: Pearson Education, Inc.

Green, A. (2014) Exploring Language Assessment and Testing. Oxon: Routledge.