

Ministry of Higher Education and
Scientific Research

University of Tikrit

College of Education for Humanities

English Department



Assessing Receptive Skills

Prof. Dr. Nagham Q.yahya

Nagyahya@tu.edu.iq

2025 – 2026

Introduction

- The purpose of assessing reading and listening is to gain insight into how well assessees can use their language abilities to understand relevant written or spoken input, often texts or recordings in the target language representing everyday life or a language learning programme.
- Users may wish to know how well assessees' proficiency in workplace or further study, understand specific difficulties that learners experience when trying to read or listen in a foreign language and help them overcome these difficulties (a diagnostic, formative purpose).
- Comprehension cannot be observed directly, so assessment designers get assessees to do something else (actions, drawing, speaking, writing) to demonstrate understanding.
- Successful performance depends on understanding input and following instructions, but assessors focus on reading or listening with understanding rather than the ability to produce language or to follow instructions, it is therefore essential that all instructions are easy for assessees and all required abilities are within the assessee's capabilities.

Obtaining the necessary evidence about reading or listening will involve:

- defining the nature of the knowledge, skills or abilities to be measured.
- choosing and specifying the kinds of texts or recordings that will be appropriate for use as input.
- deciding on the kinds of tasks that assessees with these abilities should be able to perform when they encounter these texts or recordings.

Defining reading and listening abilities:

- Building any assessment begins with careful consideration of the decision about the assessee and the knowledge, skills, or abilities relevant to that decision.
- In proficiency assessment, designers need to know which knowledge, skills, or abilities are essential in the target language use domain.
- In classroom formative assessment, the teacher must be aware of the knowledge, skills, or abilities intended to be developed by the language program.
- Designers must have an adequate understanding of the nature of the abilities they want to assess.

Types of reading:

- People use different types of reading to achieve their goals. **Expeditious reading** is quick, selective, and efficient, used when picking out a few words or phrases in a mass of information, such as finding your way in an airport, searching on the internet, or flicking through a newspaper.
- In contrast, **careful reading** is linear and not selective, where the reader follows step-by-step instructions or reads to understand all or most of the information, such as operating machinery, cooking from a recipe, or learning software.

- **Lower-level language learners** may struggle to form an overall representation of the text because they focus on decoding words and grammar.

- **Successful readers** match the type of reading to their goals and the type of text, choosing expeditious or careful reading according to the situation. Assessment designers may wish to discover whether an assessee is able to do this.

➤ **Three forms of expeditious reading:**

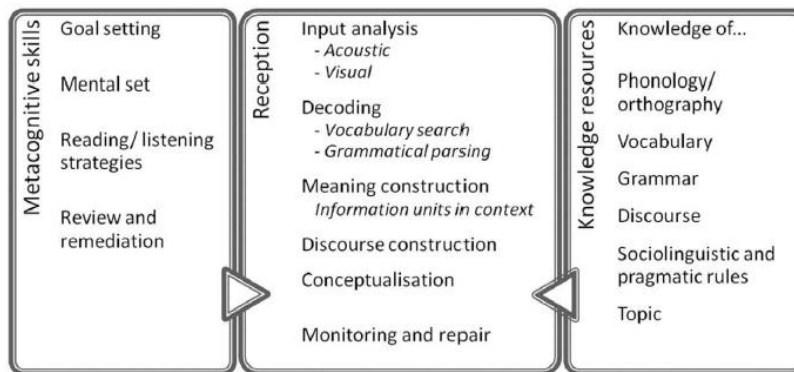
1-scanning: the reader is searching for specific words or phrases.

2-skimming or survey reading: can be useful when a reader is confronted with a lot of material, it's global the reader aims to get the overall understanding of the complex text.

3- search reading: the reader is looking for ideas relating to a certain topic.

❖ Scanning is usually located when the key information is found, but in search reading, it is a preliminary step.

Reading Process:



A model of receptive language processing based on Weir (2005a) and Field (2008)

Types of listening:

Listen for gist: understanding the main idea

listen selectively: for specific pieces of information.

listen in depth: attending to every detail.

Successful listeners like to choose the type of listening that suits the input and their goals. For example, people listen selectively to traffic news, but a learner taking a first driving lesson would listen carefully to every part of the instructor's advice.

Listening processes:

Word boundaries are not reflected in speech, and phonological effects can make it difficult for learners to recognize individual words. These effects include:

1. **Weak forms** are function words like a, the, and, but that can become almost inaudible when not stressed; different function words like are, a, and of can sound identical in continuous speech.

2. **Word transitions** occur when words occur together, sounds may be modified (assimilation) or lost (elision); final /t/ or /d/ sounds are often lost.

3. **Reduction**: The same word or phrase may sound very different when spoken carefully and when it occurs in casual speech; commonly occurring phrases are often reduced and difficult to recognize.

4. **Speaker variation**: Individuals speak in different ways (voice, rate, pausing, intonation, accent); listeners must adjust and 'tune in' to the pattern of each new voice, which is more difficult in a foreign language and with several speakers.

Choosing texts or recordings for use as input:

Finding input texts or recordings that represent an appropriate level of difficulty or challenge for assesseees must be a key consideration for the developer. A **compromise** must usually be struck between situational **authenticity** (Bachman and Palmer, 1996) and the **knowledge resources** that assesseees can be expected to bring to the task (Buck, 2001).

Indices based on **word length and sentence lengths**, such as *Flesch Reading Ease* index, **do not really capture all of the causes of comprehension problems. Long explicit sentences can be easier** for learners to understand than short elliptical sentences.

Green (2012a) suggests a wide range of measurable features that might help to predict text difficulty, but for the most part assessment designers and item writers rely on their **own judgement** in estimating the difficulty of input material.

Specifications

In building operational assessments, the selection of texts or recordings will be guided by the specifications and by their potential as sources for generating tasks and items. Designers and writers are faced with a choice. Is it best to:

- **craft new** material to use as input;
- **adapt** found source material to match the specifications;
- **try to find** material that can be used unaltered in the assessment?

For most purposes, it is said to be preferable to use texts and recordings that have been obtained from sources in the real world contexts.

However, in practice, there may **need to be a trade-off** between what is desirable from a theoretical standpoint and practical constraints on the kinds of text or recording that can be found for use in the assessment.

Much of the **reading** that language learners will need to do outside the classroom may take the form of published or publicly accessible texts.

For **listening**, if suitable recordings cannot be found for use as input to listening assessments, an alternative is to make recordings that (as far as possible) simulate real world language use.

The internet has provided item writers with a previously unimaginable wealth of potential input material in all modes from which to choose. However, for the present, there is no doubt that finding suitable sources for use in tests of comprehension remains a subtle and challenging activity. Increasing use of integrated multimedia resources can only add to the complexity of the challenge.

Assessing grammar and vocabulary

Assessments that focus on knowledge of words and phrases (i.e., traditional grammar and vocabulary tests) cover an **important but restricted** proportion of language processing.

Despite this restriction, there are many reasons why this may be worth assessing (*i.e., to keep such assessments*):

1. One is that **problems in accessing the vocabulary or grammar of a text or recording could be the source of difficulty in comprehension.**
2. Another reason is these **assessments' potential for practicality and reliability.**

But, if too much weight is given to the results of restricted grammar and vocabulary assessments, the potential for negative washback is clear.

Preparing items for assessing reading and listening

From text to task

When educators discuss the assessment of reading and listening abilities, they are talking about the comprehension of longer written texts and recordings of spoken language.

When comprehension is being assessed, the responses ought to come from an understanding of the text or recording.

The purpose for reading a text or listening to a recording in an assessment should reflect the purpose a reader would naturally have for reading similar texts in the world beyond the classroom.

Guidelines for preparing tasks

Receptive skills inputs (text or recording) should: allow you to produce (at least) the number of questions you need; be at the right level of difficulty for the assessees; be familiar and culturally suitable for all test-takers; avoid sensitive topics; not be taken from text books or other educational materials that any of the assessees may have seen before.

Writing Good Questions

Questions should: be meaningful: cover information that a reader/listener in the real world would need to extract; generally **follow the order** of the information in the text/recording; In listening tests, **allow time** for listeners to respond and refocus their attention; **include all the information needed** to find the answer; **be clear and positively worded** (especially avoid double negatives); **employ simpler language than the text/recording;**

If the above guidelines are followed, the expected **answers** should: **be unambiguous; not be recoverable from the text/recording; not be answerable without the text/recording; not depend on, be linked to other questions; not assume or depend on culturally specific knowledge.**

Answer choices (in selected response tests) should: all offer logical, grammatically possible answers to the question; all have some support from the text; all appear similar (i.e., structure and length); each be unique.

Answer choices should **NOT:** seem to offer binary choices; incorporate other options; include absolutes using words like "all", "everybody"; include qualifiers of quantity or frequency like "often", "some".

Task instructions should: be as brief as possible; explain what skills are being tested; explain the amount of time that should be spent on each task or test part; explain how many points will be awarded for each question or task; be expressed in the simplest possible language; be familiar to assessees; be accompanied by examples demonstrating how assessees should respond.

Test developers must make decisions about the most appropriate techniques to match the purpose of the assessment. The focus of the test may be on individual words, as in a test of vocabulary or grammar to understand ideas, inferences, and overall text meaning.

A problem for assessment developers is that knowledge of grammar and vocabulary and comprehension of detail are more straightforward and practical to assess than inferences, main ideas and relationships between ideas.

Trying out material

It discusses the **quality control process** in test development, which includes **reviewing** and **piloting** items. **Spaan (2007)** used a **listening test item** to show how problems can appear during **piloting**—the item was **too easy** and did not discriminate between high- and low-ability of assessees. Causes included **word repetition** and **weak distractors**. Revisions involved using **idiomatic language** and making options more **plausible**. Finally, **discussions** with assessees help improve both **assessment materials** and **learning quality**.

Scoring performance

Scoring performance mentions that **quality control procedures** for scoring **selected-response** and **short-answer items** were already discussed in Chapter 3, while methods for scoring **extended responses** will be explained in Chapter 6.

Score reporting and feedback

It explains that while external tests often delay **score reporting** and give limited **feedback**, efforts are being made to make results more **informative**, such as providing **profile scores** for different skills. In classrooms, teachers can offer **immediate feedback** and use techniques like **reflection**, **discussion**, and **peer explanation** to enhance learning. **William (2011)** suggests using **voting cards** or **mini whiteboards** to check understanding. For **extended responses**, group discussions and justifications help both teachers and learners improve **comprehension** and guide future teaching.

Standard setting

Feedback is central in **formative assessment**, but in **summative** and **proficiency assessments**, it is often limited because the main focus is on **performance standards** and **cut scores**. A **cut score** defines the level of ability needed to pass or meet a specific goal. Before setting it, the test's **validity** and **content standards** must be ensured. Some assessments use a single cut score, while others use **multiple bands** or **grades**, such as **IELTS**, which sets different required bands for subjects. Institutions may also set **local standards** based on their needs. Providers offer **sample tests** and materials to help determine fair and defensible cut scores, ensuring reliable and valid **standard setting**.