6.0 SYNOPSIS
Topic 6 focuses on ways to assess language skills and language content. It defines the types of test items used to assess language skills and language content. It also provides teachers with suggestions on ways a teacher can assess the listening, speaking, reading and writing skills in a classroom. It also discusses concepts of and differences between discrete point test, integrative test and communicative test.

6.1 LEARNING OUTCOMES
At the end of Topic 6, teachers will be able to:
• Identify and carry out the different types of assessment to assess language skills and language content
• Understand and differentiate between objective and subjective testing
• Understand and differentiate between discrete point test, integrative test and communicative test in assessing language.

6.2 FRAMEWORK OF TOPICS

CONTENT
SESSION SIX (6 hours)
6.2.1 Types of test items to assess language skills

a. Listening

Basically there are two kinds of listening tests: tests that test specific aspects of listening, like sound discrimination; and task based tests which test skills in accomplishing different types of listening tasks considered important for the students being tested. In addition to this, Brown 2010 identified four types of listening performance from which assessment could be considered.

i. **Intensive**: listening for perception of the components (phonemes, words, intonation, discourse markers, etc) of a larger stretch of language.

ii. **Responsive**: listening to a relatively short stretch of language (a greeting, question, command, comprehension check, etc.) in order to make an equally short response.

iii. **Selective**: processing stretches of discourse such as short monologues for several minutes in order to "scan" for certain information. The purpose of such performance is not necessarily to look for global or general meaning but to be able to comprehend designated information in a context of longer stretches of spoken language (such as classroom directions from a teacher, TV or radio news items, or stories). Assessment tasks in selective listening could ask students, for example, to listen for names, numbers, grammatical category, directions (in a map exercise), or certain facts and events.

iv. **Extensive**: listening to develop a top-down, global understanding of spoken language. Extensive performance ranges from listening to lengthy lectures to listening to a conversation and deriving a comprehensive message or purpose. Listening for the gist – or the main idea- and making inferences are all part of extensive listening.

b. Speaking

In the assessment of oral production, both discrete feature objective tests and integrative task-based tests are used. The first type tests such skills as pronunciation, knowledge of what language is appropriate in different situations, language required in doing different things like describing, giving directions, giving instructions, etc. The second type involves finding out if pupils can perform different tasks using spoken language that is appropriate for the purpose and the context. Task-based activities involve describing scenes shown in a picture, participating in a discussion about a given topic, narrating a story, etc. As in the

1. Imitative. At one end of a continuum of types of speaking performance is the ability to imitate a word or phrase or possibly a sentence. Although this is a purely phonetic level of oral production, a number of prosodic (intonation, rhythm, etc.), lexical, and grammatical properties of language may be included in the performance criteria. We are interested only in what is traditionally labelled “pronunciation”; no inference are made about the test-takers ability to understand or convey meaning or to participate in an interactive conversation. The only role of listening here is in the short-term storage of a prompt, just long enough to allow the speaker to retain the short stretch of language that must be imitated.

2. Intensive. The production of short stretches of oral language designed to demonstrate competence in a narrow band of grammatical, phrasal, lexical, or phonological relationships. Examples of intensive assessment tasks include directed response tasks (requests for specific production of speech), reading aloud, sentence and dialogue completion, limited picture-cued tasks including simple sentences, and translation up to the simple sentence level.

3. Responsive. Responsive assessment tasks include interaction and test comprehension but at somewhat limited level of very short conversation, standard greetings, and small talk, simple requests and comments, etc. The stimulus is almost always a spoken prompt (to preserve authenticity) with one or two follow-up questions or retorts:
   A. Liza : Excuse me, do you have the time?  
      Don : Yeah. Six-fifteen.  
   B. Jo : What is the most urgent social problem today?  
      Sue : I would say bullying.  
   C. Lan : Hey, Shan, how’s it going?  
      Shan : Not bad, and yourself?  
      Lan : I’m good.  
      Shan : Cool. Okay gotta go.

4. Interactive. The difference between responsive and interactive speaking is in the length and complexity of the interaction, which sometimes includes multiple
exchanges and/or multiple participants. Interaction can be broken down into two types: (a) transactional language, which has the purpose of exchanging specific information, and (b) interpersonal exchanges, which have the purpose of maintaining social relationships. (In the three dialogues cited above, A and B are transactional, and C is interpersonal).

5. Extensive (monologue). Extensive oral production tasks include speeches, oral presentations, and storytelling, during which the opportunity for oral interaction from listeners is either highly limited (perhaps to nonverbal responses) or ruled out together. Language style is more deliberative (planning is involved) and formal for extensive tasks. In can include informal monologue such as casually delivered speech (e.g., recalling a vacation in the mountains, conveying recipes, recounting the plot of a novel or movie).

c. Reading
Cohen (1994), discussed various types of reading and meaning assessed. He describes skimming and scanning as two different types of reading. In the first, a respondent is given a lengthy passage and is required to inspect it rapidly (skim) or read to locate specific information (scan) within a short period of time. He also discusses receptive reading or intensive reading which refers to “a form of reading aimed at discovering exactly what the author seeks to convey” (p. 218). This is the most common form of reading especially in test or assessment conditions. Another type of reading is to read responsively where respondents are expected to respond to some point in a reading text through writing or by answering questions.

A reading text can also convey various kinds of meaning and reading involves the interpretation or comprehension of these meanings. First, grammatical meaning are meanings that are expressed through linguistic structures such as complex and simple sentences and the correct interpretation of those structures. A second meaning is informational meaning which refers largely to the concept or messages contained in the text. Respondents may be required to comprehend merely the information or content of the passage and this may be assessed through various means such as summary and précis writing. Compared to grammatical or syntactic meaning, informational meaning requires a more general understanding of a text rather than having to pay close attention to the linguistic structure of sentences. A third meaning contained in many texts is discourse meaning. This refers to the
perception of rhetorical functions conveyed by the text. One typical function is discourse marking which adds cohesiveness to a text. These words, such as unless, however, thus, therefore etc., are crucial to the correct interpretation of a text and students may be assessed on their ability to understand the discoursal meaning that they bring in the passage. Finally, a fourth meaning which may also be an object of assessment in a reading test is the meaning conveyed by the writer’s tone. The writer’s tone – whether it is cynical, sarcastic, sad or etc.- is important in reading comprehension but may be quite difficult to identify, especially by less proficient learners. Nevertheless, there can be many situations where the reader is completely wrong in comprehending a text simply because he has failed to perceive the correct tone of the author.

d. Writing

Brown (2004), identifies three different genres of writing which are academic writing, job-related writing and personal writing, each of which can be expanded to include many different examples. Fiction, for example, may be considered as personal writing according to Brown's taxonomy. Brown (2010) identified four categories of written performance that capture the range of written production which can be used to assess writing skill.

1. Imitative. To produce written language, the learner must attain the skills in the fundamental, basic tasks of writing letters, words, punctuation, and brief sentences. This category includes the ability to spell correctly and to perceive phoneme-grapheme correspondences in the English spelling system. At this stage the learners are trying to master the mechanics of writing. Form is the primary focus while context and meaning are of secondary concern.

2. Intensive (controlled). Beyond the fundamentals of imitative writing are skills in producing appropriate vocabulary within a context, collocation and idioms, and correct grammatical features up to the length of a sentence. Meaning and context are important in determining correctness and appropriateness but most assessment tasks are more concerned with a focus on form and are rather strictly controlled by the test design.

3. Responsive. Assessment tasks require learners to perform at a limited discourse level, connecting sentences into a paragraph and creating a logically connected sequence of two or three paragraphs. Tasks relate to pedagogical directives, lists of criteria, outlines, and other guidelines. Genres of writing include brief narratives and descriptions, short reports, lab reports,
summaries, brief responses to reading, and interpretations of charts and
graphs. Form-focused attention is mostly at the discourse level, with a strong
emphasis on context and meaning.

4. Extensive. Extensive writing implies successful management of all
the processes and strategies of writing for all purposes, up to the length of an
essay, a term paper, a major research project report, or even a thesis. Focus
is on achieving a purpose, organizing and developing ideas logically, using
details to support or illustrate ideas, demonstrating syntactic and lexical
variety, and in many cases, engaging in the process of multiple drafts to
achieve a final product. Focus on grammatical form is limited to occasional
editing and proofreading of a draft.

6.2.2 Objective and Subjective test
Tests have been categorized in many different ways. The most familiar terms
regarding tests are the objective and subjective tests. We normally associate
objective tests with multiple choice question type tests and subjective tests with
essays. However, to be more accurate we will consider how the test is graded.
Objective tests are tests that are graded objectively while subjective tests are
thought to involve subjectivity in grading.

There are many examples of each type of test. Objective type tests include the
multiple choice test, true false items and matching items because each of these
are graded objectively. In these examples of objective tests, there is only one
correct response and the grader does not need to subjectively assess the
response.

Examples of the subjective test include essays and short answer questions.
However some other types of common tests such as the dictation test, filling in
the blank type tests, as well as interviews and role plays can be considered
subjective and objective type tests where they fall on some sort of continuum
where some tests are more objective than others. As such, some of these tests
would fall closer to one end of the continuum or the other.

Two other terms, select type tests and supply type tests are related terms when
we think of objective and subjective tests. In most cases, objective tests are
similar to select type tests where students are expected to select or choose the
answer from a list of options. Just as a multiple choice question test is an
objective type test, it can also be considered a select type test. Similarly, tests involving essay type questions are supply type as the students are expected to supply the answer through their essay. How then would you classify a fill in the blank type test? Definitely for this type of test, the students need to supply the answer, but what is supplied is merely a single word or a short phrase which differs tremendously from an essay. It may therefore be helpful to once again consider a continuum with supply type and select type items at each end of the continuum respectively.

It is possible to now combine both continua as shown in Figure 6.1 with the two different test formats placed within the two continua:

![Figure 6.1: Continua for different types of test formats](image)

It is not by accident that we find there are few, if any, test formats that are either supply type and objective or select type and subjective. Select type tests tend to be objective while supply type tests tend to be subjective.

In addition to the above, Brown and Hudson (1998), have also suggested three broad categories to differentiate tests according to how students are expected to respond. These categories are the selected response tests, the constructed response tests, and the personal response tests. Examples of each of these types of tests are given in Table 6.1.

**Table 6.1: Types of Tests According to Students’ Expected Response**

<table>
<thead>
<tr>
<th>Selected response</th>
<th>Constructed response</th>
<th>Personal response</th>
</tr>
</thead>
<tbody>
<tr>
<td>True false</td>
<td>Fill-in</td>
<td>Conferences</td>
</tr>
<tr>
<td>Matching</td>
<td>Short answer</td>
<td>Portfolios</td>
</tr>
<tr>
<td>Multiple choice</td>
<td>Performance test</td>
<td>Self and peer assessments</td>
</tr>
</tbody>
</table>
Selected response assessments, according to Brown and Hudson (1998), are assessment procedures in which “students typically do not create any language” but rather “select the answer from a given list” (p. 658). Constructed response assessment procedures require students to “produce language by writing, speaking, or doing something else” (p. 660). Personal response assessments, on the other hand, require students to produce language but also allows each students’ response to be different from one another and for students to “communicate what they want to communicate” (p. 663). These three types of tests, categorised according to how students respond, are useful when we wish to determine what students need to do when they attempt to answer test questions.

6.2.3 Types of test items to assess language content

a. Discrete Point Test and Integrative Test

Language tests may also be categorised as either discrete point or integrative. Discrete point tests examine one element at a time. Integrative tests, on the other hand, “requires the candidate to combine many language elements in the completion of a task” (Hughes, 1989: 16). It is a simultaneous measure of knowledge and ability of a variety of language features, modes, or skills.

A multiple choice type test is usually cited as an example of a discrete point test while essays are commonly regarded as the epitome of integrative tests. However, both the discrete point test and the integrative test are a matter of degree. A test may be more discrete point than another and similarly a test may be more integrative than another. Perhaps the more important aspect is to be aware of the discrete point or integrative nature of a test as we must be careful of what we believe the test measures.

This brings us to the question of how discrete point is a multiple choice question type item? While it is definitely more discrete point than an essay, it may still require more than just one skill or ability in order to complete. Let’s say you are interested in testing a student’s knowledge of the relative pronoun and decide to do so by using a multiple choice test item. If he fails to answer this test item correctly, would you conclude that the student has problems with the relative pronoun? The answer may not be as straight forward as it seems. The test is presented in textual form and therefore requires the student to read. As such, even the multiple choice test item involves some integration of language skills as
this example shows, where in addition to the grammatical knowledge of relative pronouns, the student must also be able to read and understand the question.

Perhaps a clearer way of viewing the distinction between the discrete point and the integrative test is to examine the perspective each takes toward language. In the discrete point test, language is seen to be made up of smaller units and it may be possible to test language by testing each unit at a time. Testing knowledge of the relative pronoun, for example, is certainly assessing the students on a particular unit of language and not on the language as a whole. In an integrative test, on the other hand, the perspective of language is that of an integrated whole which cannot be broken up into smaller units or elements. Hence, the testing of language should maintain the integrity or wholeness of the language.

b. Communication Test

As language teaching has emphasised the importance of communication through the communicative approach, it is not surprising that communicative tests have also been given prominence. A communicative emphasis in testing involves many aspects, two of which revolve around communicative elements in tests and meaningful content. Both these aspects are briefly addressed in the following subsections:

Integrating Communicative Elements into Examinations

Alderson and Banerjee (2002), report on various studies that seem to point to the difficulty in achieving authenticity in tests. They cite Spence-Brown (2001) who posits that “the very act of assessment changes the nature of a potentially authentic task and compromises authenticity” and that “authenticity must be related to the implementation of an activity, not to its design” (p. 99). In her study, students were required to interview native speakers outside the classroom and submit a tape-recording of the interview. While this activity seems quite authentic, the students were observed to prepare for the interview by “rehearsing the interview, editing the results, and engaging in spontaneous, but flawed discourse” (Alderson & Banerjee, 2002: 99), all of which are inauthentic when viewed in terms of real life situations. Alderson himself argues that because candidates in language tests are not interested in communicating but to display their language abilities, the test situation is a communicative event in itself and therefore cannot be used to replicate any real world event (p. 98).
Chalhoub-Deville (2003), argues for tests that take context into consideration. She believes that there should be a “shift in focus of our measurement from traditional examinations of the construct in terms of response consistency, to investigations that systematically explore inconsistent (which does not mean random) performances across contexts” (p. 378). In the future, besides context, tests will also need to integrate elements of communication such as topic initiation, topic maintenance, and topic change in order for the test to become more authentic and realistic. Due to issues of practicality, involving especially the amount of time and extent of organisation to allow for such communicative elements to emerge, it will not be an easy task to achieve.

The idea of bringing communicative elements into the language test is not a new one. In his review of communicative tests, Fulcher (2000), notes the descriptors of a communicative test as suggested by several theorists. The three principles of communicative tests that he highlights are that communicative tests:

- involve performance;
- are authentic; and
- are scored on real-life outcomes.

In short, the kinds of tests that we should expect more of in the future will be communicative tests in which candidates actually have to produce the language in an interactive setting involving some degree of unpredictability which is typical of any language interaction situation. These tests would also take the communicative purpose of the interaction into consideration and require the student to interact with language that is actual and unsimplified for the learner. Fulcher finally points out that in a communicative test, “the only real criterion of success … is the behavioural outcome, or whether the learner was able to achieve the intended communicative effect” (p. 493). It is obvious from this description that the communicative test may not be so easily developed and implemented. Practical reasons may hinder some of the demands listed. Nevertheless, a solution to this problem has to be found in the near future in order to have valid language that are purposeful and can stimulate positive washback in teaching and learning.
Exercise 1

1. In your opinion and based on your teaching experience, how would you conduct the testing of reading, writing and speaking skills of your own students? What are the methods that you employ? Share this with your classmates and exchange ideas.

2. Describe three different types of writing performance as suggested by Brown (2004) and relate their relationship to academic writing, job related writing and personal writing.