

УЗБЕКИСТОН РЕСПУБЛИКАСИ ХАЛҚ ТАЪЛИМИ

ВАЗИРЛИГИ

**АЖИНИЁЗ НОМИДАГИ НУКУС ДАВЛАТ
ПЕДАГОГИКА ИНСТИТУТИ**

ЧЕТ ТИЛЛАРИ ФАКУЛЬТЕТИ

ИНГЛИЗ ТИЛИ ВА АДАБИЁТИ КАФЕДРАСИ

Тилди тексеру хам бахалау ?Language testing ?

ФАНИДАН

МАЪРУЗА МАТНЛАРИ

ТУЗУВЧИ:

КОЙШЕКЕНОВА Т.К.

LECTURE 1. INTRODUCTION. PRINCIPLES OF LANGUAGE ASSESSMENT

Plan:

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

Glossary: *practicality, reliability, validity, content-related evidence, criterion-related evidence, construct-related evidence, authenticity, washback, test taking strategies.*

Introduction

This lecture explores how principles of language assessment can and should be applied to formal tests, but with the ultimate recognition that these principles also apply to assessments of all kinds. How do you know if a test is effective? For the most part that question can be answered by responding to such questions as:

Can it be given within appropriate administrative constraints?

Is it dependable?

Does it accurately measure what you want it to measure?

These and other questions help to identify 5 cardinal criteria for “testing a test”:

practicality, reliability, validity, authenticity and wash-back. We will look at each one.

1.1. Practicality.

An effective test is practical. This means that it:

- 1 is not expensive;
- 2 stays within appropriate time constraints;
- 3 is easy to administer;
- 4 has a scoring or evaluation procedure that is specific and time-efficient.

A test that is expensive is impractical. Practicality is determined by the teacher's and students' time constraints, costs, administrative details, and to some extent by what occurs before and after the test. To determine if a test practical for your needs, you may want to use the checklist below.

Practicality checklist

- 1. Are administrative details clearly established before the test?
- 2. Can students complete the test reasonably within the set time frame?
- 3. Can the test be administered smoothly, without procedural "glitches"?
- 4. Are all materials and equipment ready?
- 5. Is the cost of the test within budgeted limits?
- 6. Is the scoring/evaluation system feasible in the teacher's time frame?
- 7. Are methods for reporting results determined in advance?

1.2. Reliability.

A reliable test is consistent and dependable. If you give the same test to the same student or matched students on two different occasions, the test should yield similar results. The issue of reliability of a test may best be addressed by considering a number of factors that may contribute to the unreliability of a test. You should consider the following possibilities: fluctuations in the student, in scoring, in test administration and in the test itself.

Student-related reliability. The most common learner-related issue in reliability is caused by temporary illness, fatigue (fə'ti:g - уст алост ь), a "bad day", anxiety, and other physical or psychological factors, which may make an "observed score" deviate from one's "true" score. Such factors as a test-taker's "test-wiseness" or strategies for efficient test taking are also included in this category.

Rater reliability. Human error, subjectivity and bias (baiəs - уклон) may enter into the scoring process. *Inter-rater reliability* occurs when two or more scorers yield inconsistent scores of the same test, possibly for lack of attention to scoring criteria, inexperience and inattention.

Intra-rater reliability is a common occurrence for classroom teachers because of unclear scoring criteria, fatigue, bias toward particular "good" and "bad" students or simple carelessness. J.D.Brown mentioned: "The careful specification of an analytical scoring instrument can increase rater reliability".

Test administration reliability. Unreliability may result from the conditions in which the test is administered. Sometimes we witness the administration of a test of aural comprehension because of street noise outside the building, students sitting next to windows can't hear the tape accurately. Other sources of unreliability are found in photocopying variations, the amount of light in different

parts of the room, variations in temperature and even the conditions of desks and chairs.

Test reliability. Sometimes the nature of the test itself can cause measurement errors. If a test is too long, test-takers may become fatigued by the time they reach the later items and respond incorrectly. Timed tests may discriminate against students who don't perform well on a test with a time limit. We all know people who "know" the material perfectly but who are adversely affected by the presence of a clock ticking away.

Thus, reliability applies to both the test and the teacher. Test and test administration reliability can be achieved by making sure that all students receive the same quality of input, whether written or auditory. Part of achieving test reliability depends on the physical context, that:

- 1 every student has a cleanly photocopied test sheet;
- 2 sound amplification is clearly audible to everyone in the room;
- 3 video input is visible to all;
- 4 lighting, temperature, noise and other conditions are equal for all students;
- 5 objective scoring procedures leave little debate about correctness of an answer.

1.3. Validity.

The most complex criterion of an effective test and the most important principle is validity. It is the extent to which inferences made from assessment results are meaningful and useful in terms of the purpose of the assessment. We will look at 5 types of evidence below.

1. Content-related evidence. This form of evidence relates to the content of the test. If a test samples the subject matter about which conclusions are to be drawn, and if it requires the test-taker to perform the behavior that is being measured, it can claim content-related evidence of validity. You can identify it if you can clearly define the achievement that you are measuring. Consider the following quiz on English articles for learners.

2. Criterion-related evidence is a second form of evidence of the test validity. It is the extent to which the "criterion" of the test has been reached. Classroom-based assessment with teacher-designed tests fits the concept of criterion-referenced assessment. In such tests, specified classroom objectives are measured, and implied predetermined levels of performance are expected to be reached.

3. Construct-related evidence is a third kind of evidence that support validity. A construct is any theory, hypothesis or model that attempts to explain observed phenomena in our universe of perceptions. "Proficiency" and "communicative competence" are linguistic constructs, "self-esteem" and "motivation" are psychological constructs. Construct validity is a major issue in validating large-scale standardized tests of proficiency. Because such tests must adhere to the principle of practicality.

4. Consequential validity encompasses all the consequences of a test including such considerations as its accuracy in measuring intended criteria, its impact on the preparation of test-takers, its effect on the learner, and the social consequences of a test's interpretation and use.

5. Face validity means that the students perceive the test to be valid. Face validity will be high if learners encounter:

- 1 a well-constructed, expected format with familiar tasks;
- 2 a test that is clearly doable within the allotted time limit;
- 3 items that are clear and uncomplicated;
- 4 directions that are crystal clear;
- 5 tasks that relate to their course work;
- 6 a difficulty level that presents a reasonable challenge.

As already noted above, content-validity is a very important component in achieving face validity.

In evaluating a classroom test teachers should consider following test-taking strategies:

Before the Test

1. Give students all the information you can about the test: Exactly what the test cover? Which topics will be the most important? What kind of items will be on it? How long will it be?
2. Encourage students to do a systematic review of material. For example they should skim the textbook and other material, outline major points and write down examples.
3. Give them practice tests or exercises, if available.
4. Facilitate formation of a study group, if possible.
5. Caution students to get a good night's rest before the test.
6. Remind students to get to the classroom early.

During the Test

1. After the test is distributed, tell students to look over the whole test quickly in order to get a good grasp of its different parts.
2. Remind them to mentally figure out how much time they will need for each part.
3. Advise them to concentrate as carefully as possible.
4. Warn students a few minutes before the end of the class period so that they can finish on time, proofread their answers, and catch careless mistakes.

After the Test

1. When you return the test, include feedback on specific things the student did well, what he or she did not do well, and, if possible, the reasons for your comments.
2. Advise students to pay careful attention in class to whatever you say about the test results.
3. Encourage questions from students.
4. Advise students to pay special attention in the future to points on which they are weak.

1.4. Authenticity.

A fourth principle of language testing is authenticity, a concept that is a little slippery to define, especially within the art and science of evaluating and designing tests. Bachman and Palmer define authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a target language task. In a test, authenticity may be present in the following ways:

- 1 The language in the test is as natural as possible.
- 2 Items are contextualized rather than isolated.
- 3 Topics are meaningful for the learner.
- 4 Some thematic organization to items is provided, such as through a story

line or episode.

5 Tasks represent real-world tasks.

The authenticity of test tasks in recent years has increased noticeably. Two or three decades ago, unconnected, boring, contrived items were accepted as a necessary component of testing. Things have changed. It was once assumed that large-scale testing could not include performance of the productive skills and stay within budgetary constraints, but now many such tests offer speaking and writing components. Reading passages are selected from real-life sources that test-takers are likely to have encountered or will encounter. Listening comprehension sections feature natural language with hesitations, white noise and interruptions. More and more tests offer items that are “episodic” in that they are sequences to form meaningful units, paragraphs or stories.

1.5. Washback

A facet of consequential validity, discussed above is the effect of testing on teaching and learning, otherwise known among language-testing specialists as washback. In large-scale assessment, washback refers to the effects, the tests have on instruction in terms of how students prepare for the test.

Another form of washback that occurs more in classroom assessment is the information that “washes back” to students in the form of useful diagnoses of strengths and weaknesses. Washback also includes the effects of an assessment on teaching and learning prior to the assessment itself, that is, on preparation for the assessment.

Finally, washback also implies that students have ready access to you to discuss the feedback and evaluation you have given. While you almost certainly have known teachers with whom you wouldn’t dare argue about a grade, an interactive, cooperative and collaborative classroom can promote an atmosphere of dialogue between students and teachers regarding evaluate judgments. For learning to continue, students need to have a chance to feed back on your feedback, to seek clarification of any issues that are fuzzy, and to set new and appropriate goals for themselves for the days and weeks ahead.

Thus, the 5 principles of practicality, reliability, validity, authenticity and washback go a long way toward providing useful guidelines for both evaluating an existing assessment procedure and designing one on your own. Quizzes, tests, final exams and proficiency tests can all be scrutinized through these 5 lenses.

Suggested Literature:

LECTURE 2. DESIGNING CLASSROOM LANGUAGE TESTS

Plan:

1. Test types.
2. Some practical steps to test construction.
3. Scoring, grading and giving feedback.

Glossary: *language aptitude tests, language proficiency tests, placement tests,*

diagnostic tests, achievement tests, test construction, test specification, scoring, grading, giving feedback, multiple-choice test items.

In this lecture, you will draw on the foundations and tools to begin the process of designing tests or revising existing tests. To start that process, you need to ask some critical questions:

What is the purpose of the test?

What are the objectives of the test?

How will the test specifications reflect both the purpose and objectives?

How will the test tasks be selected and the separate items arranged?

What kind of scoring, grading and feedback is expected?

These 5 questions should form the basis of your approach to designing tests for your classroom.

2.1. Test types.

The first task you will face in designing a test for your students is to determine the purpose for the test. Defining your purpose will help you choose the right kind of test, and it will also help you to focus on the specific objectives of the test. We will look first at two test types that you will not have many opportunities to create as a classroom teacher – **language aptitude tests** and **language proficiency tests** – and three types that you will certainly need to create – **placement tests**, **diagnostic tests** and **achievement tests**.

Language Aptitude Tests. A language aptitude test is designed to measure capacity or general ability to learn a foreign language and ultimate success in that undertaking. Language aptitude tests are ostensibly designed to apply to the classroom learning of any language. Two standardized aptitude tests have been used in the USA: the Modern Language Aptitude Test (MLAT-1958) and Pimsleur Language Aptitude Battery (PLAB-1966). Both are English language tests and require students to perform a number of language-related tasks. Because of limitation LAP are seldom used today.

Proficiency Tests. A proficiency test tests overall ability. Proficiency Tests have traditionally consisted of standardized multiple-choice items on grammar, vocabulary, reading comprehension and aural comprehension. Sometimes a sample of writing is added, and more recent tests also include oral production performance. A typical example of a proficiency test is the Test of English as a Foreign Language (TOEFL) produced by Educational Testing Service and IELTS (The International English language testing system). In the case of some proficiency tests, 'proficient' means having sufficient command of the language for the particular purpose.

Placement Tests. Placement tests are intended to provide information that will help to place students at the stage of the teaching programme most appropriate to their abilities. Typically they are used to assign students to classes at different level. A placement test includes a sampling of the material to be covered in the various courses in a curriculum.

Diagnostic Tests. A diagnostic test is designed to diagnose specified aspects of a language. Such kind of tests are used to identify learners' strengths and weaknesses. A test in pronunciation, for example, might diagnose the phonological features of English that are difficult for learners and should therefore become part of a curriculum. Usually, such tests offer a checklist of features for the administrator (often the teacher) to use in pinpointing difficulties. A writing diagnostic would elicit a writing sample from students that would allow the teacher to identify those rhetorical and linguistic features on which the course needed to focus special attention.

A typical diagnostic test of oral production was created by Clifford Prator (1972) to accompany a manual of English

pronunciation. Test-takers are directed to read a 150-word passage while they are tape-recorded. The test administrator then refers to an inventory of phonological items for analyzing a learner's production. After multiple listening, the administrator produces a checklist of errors in five separate categories, each of which has several subcategories. The main categories include:

1. stress and rhythm;
2. intonation;
3. vowels,
4. consonants, and
5. other factors.

Achievement Tests. An achievement test is related directly to classroom lessons, units, or even a total curriculum. Achievement tests are (or should be) limited to particular material addressed in a curriculum within a particular time frame and are offered after a course has focused on the objectives in question.

The specifications for an achievement test should be determined by:

- the objectives of the lesson, unit, or course being assessed;
- the relative importance (or weight) assigned to each objective;
- the tasks employed in classroom lessons during the unit of time;
- practicality issues, such as the time frame for the test and turnaround time;
- the extent to which the test structure lends itself to formative washback.

Achievement tests range from five- or ten-minute quizzes to three-hour final examinations, with an almost infinite variety of item types and formats. Here is the outline for a midterm examination offered at the high-intermediate level of an intensive English program in the United States.

Section A. Vocabulary

Part 1 (5 items): match words and definitions

Part 2 (5 items): use the word in a sentence

Section B. Grammar

(10 sentences): error detection (underline or circle the error)

Section C. Reading comprehension

(2 one-paragraph passages): four short-answer items for each

Section D. Writing

respond to a two-paragraph article on Native American culture

2.2. Some practical steps to test construction.

New and innovative testing formats take a lot of effort to design and a long time to refine through trial and error. Traditional testing techniques can conform to the spirit of an interactive, communicative language curriculum. Your best tack as a new teacher is to work within the guidelines of accepted, known, traditional testing techniques. Slowly, with experience, you can get bolder in your attempts. In that spirit, then, let us consider some practical steps in constructing classroom tests.

1. Assessing Clear, Unambiguous Objectives. In addition to knowing the purpose of the test you're creating, you need to know as specifically as possible what it is you want to test. **You should** begin by taking a careful look at everything that you think your students should "know" or be able to "do" based on the material that the students are responsible for. In other words, examine the **objectives** for the unit you are testing.

Selected objectives for a unit in a low-intermediate integrated-skills course

Form-focused objectives (listening and speaking)

Students will

1. recognize and produce tag questions, with the correct grammatical form and final intonation pattern, in simple social conversations.
2. recognize and produce *wh*-information questions with correct final intonation pattern.

Communication skills (speaking)

Students will

3. state completed actions and events in a social conversation.
4. ask for confirmation in a social conversation.
5. give opinions about an event in a social conversation.
6. produce language with contextually appropriate intonation, stress, and rhythm.

Reading skills (simple essay or story)

Students will

7. recognize irregular past tense of selected verbs in a story or essay.

Writing skills (simple essay or story)

Students will

8. write a one-paragraph story about a simple event in the past.
9. use conjunctions *so* and *because* in a statement of opinion.

2. Drawing Up Test Specifications. Test specifications for classroom use can be a simple and practical outline of your test.

Speaking (5 minutes per person, previous day)

Format: oral interview, T and S

Task: T asks questions of S (objectives 3, 5; emphasis on 6)

Listening (10 minutes)

Format: T makes audiotape in advance, with one other voice on it

- Tasks:*
- a. 5 minimal pair items, multiple-choice (objective 1)
 - b. 5 interpretation items, multiple-choice (objective 2)

Reading (10 minutes)

Format: cloze test items (10 total) in a story line

Tasks: fill-in-the-blanks (objective 7)

Writing (10 minutes)

Format: prompt for a topic: why I liked/didn't like a recent TV sitcom

Task: writing a short opinion paragraph (objective 9)

These informal, classroom-oriented specifications give you an indication of

- the topics (objectives) you will cover;
- the implied elicitation and response formats for items;
- the number of items in each section;
- the time to be allocated for each.

3. Devising Test Tasks. Your oral interview comes first, and so you draft questions to conform to the accepted pattern of oral interviews. You begin and end with non-scored items designed to set students at ease, and then sandwich between them items intended to test the objective (*level check*) and a little beyond (*probe*).

A. Warm-up: questions and comments.

B. Level-check questions.

1. Tell me about what you did last weekend.
2. Tell me about an interesting trip you took in the last year.
3. How did you like the TV show we saw this week?

C. Probe.

1. What is your opinion about? (news event)
2. How do you feel about _____ ? (another news event)

D. Wind-down: comments and reassurance

In revising your draft, you will want to ask yourself some important questions:

1. Are the directions to each section absolutely clear?
2. Is there an example item for each section?
3. Does each item measure a specified objective?
4. Is each item stated in clear, simple language?
5. Does each multiple-choice item have appropriate distractors: that is, are the wrong items clearly wrong and yet sufficiently "alluring" that they aren't ridiculously easy?
6. Is the difficulty of each item appropriate for your students?

4. Designing Multiple-Choice Test Items. Multiple-choice items are extremely difficult to design correctly. Hughes (2003, pp. 76-78) cautions against a number of weaknesses of multiple-choice items:

- The technique tests only recognition knowledge.
- Guessing may have a considerable effect on test scores.
- The technique severely restricts what can be tested.
- It is very difficult to write successful items.
- Washback may be harmful.
- Cheating may be facilitated.

2.3. Scoring, grading and giving feedback.

Scoring.

As you design a classroom test, you must consider how the test will be scored and graded. Your scoring plan reflects the relative weight that you place on each section and items in each section. How do you assign scoring to the various components of this test?

Because oral production is a driving force in your overall objectives, you decide to place more weight on the speaking (oral interview) section than on the other three sections. Five minutes is actually a long time to spend in a one-on-one situation with a student, and some significant information can be extracted from such a session. You therefore designate 40 percent of the grade to the oral interview. You consider the listening and reading sections to be equally important, but each of them, especially in this multiple-choice format, is of less consequence than the oral interview. So you give each of them a 20 percent weight. That leaves 20 percent for the writing section, which seems about right to you given the time and focus on writing in this unit of the course.

Your next task is to assign scoring for each item. This may take a little numerical common sense, but it doesn't require a degree in math. To make matters simple, you decide to have a 100-point test in which

- 1 the listening and reading items are each worth 2 points.
- 2 the oral interview will yield four scores ranging from 5 to 1, reflecting fluency, prosodic features, accuracy of the target grammatical objectives, and discourse appropriateness.
- 3 the writing sample has two scores: one for grammar/mechanics and one for overall effectiveness of the message, each ranging from 5 to 1. Again, *to* achieve the correct weight for writing, you will double each score and add them, so the possible total is 20 points.

Here are your decisions about scoring your test:

	Percent of Total Grade	Possible Correct
Oral Interview	40%	4 scores, 5 to 1 range $\times 2 = 8$
Listening	20%	10 items @ 2 points each = 20
Reading	20%	10 items @ 2 points each = 20
Writing	20%	2 scores, 5 to 1 range $\times 2 = 4$
Total		100

Grading

Your first thought might be that assigning grades to student performance on this test would be easy: just give an "A" for 90-100 percent, a "B" for 80-89 percent, and so on. Not so fast! Grading is such a thorny issue. How you assign letter grades to this test is a product of

- 1 the country, culture and context of this English classroom;
- 2 institutional expectations (most of them unwritten).
- 3 explicit and implicit definitions of grades that you have set forth;
- 4 the relationship you have established with this class;
- 5 student expectations that have been engendered in previous tests and quizzes in the class.

We will have full information about grading during LECTURE 9.

Giving Feedback

A lecture on scoring and grading would not be complete without some consideration of the forms in which you will offer feedback to your students, feedback that you want to become beneficial washback. You might choose to return the test to the student with one of, or a combination of, any of the possibilities below:

1. a letter grade.
2. a total score.
3. four sub-scores (speaking, listening, reading, writing)
4. for the listening and reading sections.
 - a. an indication of correct/incorrect responses
 - b. marginal comments
5. for the oral interview.
 - c. scores for each element being rated
 - d. a checklist of areas needing work
 - e. oral feedback after the interview
 - f. a post-interview conference to go over the results
6. on the essay.
 - g. scores for each element being rated
 - h. a checklist of areas needing work
 - i. marginal and end-of-essay comments, suggestions
 - j. a post-test conference to go over work
 - k. a self-assessment
7. on all or selected parts of the test, peer checking of results.
8. a whole-class discussion of results of the test.
9. individual conferences with each student to review the whole test.

In this lecture, guidelines and tools were provided to enable you to address the five questions posed at the outset: (1) how to determine the purpose or criterion of the test, (2) how to state objectives, (3) how to design specifications, (4) how to select and arrange test tasks, including evaluating those tasks with item indices, and (5) how to ensure appropriate washback to the student. This five-part template can serve as a pattern as you design classroom tests.

SUGGESTED LITERATURE

1. Carroll, John B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In Thomas S. Parry & Charles W. Stansfield (Eds.), *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice Hall Regents.
2. Brown, James Dean. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

3. Gronlund, Norman E. (1998). *Assessment of student achievement*. Sixth Edition. Boston: Allyn and Bacon.

LECTURE 3. ASSESSING LISTENING AND SPEAKING

Plan:

Basic types of listening

Designing Assessment Tasks: Intensive Listening.

Designing Assessment Tasks: Responsive Listening.

Designing Assessment Tasks: Selective Listening.

Designing Assessment Tasks: Extensive Listening.

Glossary: *importance of listening, basic types of listening, designing assessment tasks, intensive listening, responsive listening, selective listening, extensive listening, note-taking, editing, retelling, interpretive tasks.*

Introduction. The Importance of Listening

Listening has often played second fiddle to its counterpart, speaking. In the standardized testing industry, a number of separate oral production tests are available, but it is rare to find just a listening test. One reason for this emphasis is that listening is often implied as a component of speaking. How could you speak a language without also listening? In addition, the overtly observable nature of speaking renders it more empirically measurable than listening.

Every teacher of language knows that one's oral production ability is only as good as one's listening comprehension ability.

We therefore need to pay close attention to listening as a mode of performance for assessment in the classroom. In this lecture, we will begin with basic principles and types of listening, then move to a survey of tasks that can be used to assess listening.

1. Basic types of listening

As with all effective tests, designing appropriate assessment tasks in listening begins with the specification of objectives, or criteria. Those objectives may be classified in terms of several types of listening performance. Think about what you do when you listen. Literally in nanoseconds, the following processes flash through your brain:

1. You recognize speech sounds and hold a temporary "imprint" of them in short-term memory.
2. You simultaneously determine the type of speech event (monologue, interpersonal dialogue, transactional dialogue) that is being processed and attend to its context (who the speaker is, location, purpose) and the content of the message.

3. You use (bottom-up) linguistic decoding skills and/or (top-down) background schemata to bring a plausible interpretation to the message, and assign a *literal* and *intended meaning* to the utterance.
4. In most cases, you delete the exact linguistic form in which the message was originally received in favor of conceptually retaining important or relevant information in long-term memory.

Each of these stages represents a potential assessment objective:

1. comprehending of surface structure elements such as phonemes, words, intonation, or a grammatical category.
2. understanding of pragmatic context.
3. determining meaning of auditory input.
4. developing the gist, a global or comprehensive understanding.

From these stages we can derive four commonly identified types of listening performance, each of which comprises a category within which to consider assessment tasks and procedures:

1. **Intensive.** Listening for perception of the components (phonemes, words, intonation, discourse markers, etc.) of a larger stretch of language.
2. **Responsive.** Listening to a relatively short stretch of language (a greeting, question, command, comprehension check, etc.) in order to make an equally short response.
3. **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 meanings, 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, a grammatical category, directions (in a map exercise), or certain facts and events.
4. **Extensive.** Listening to develop a top-down, global understanding of spoken language. Listening for the gist, for the main idea, and making inferences are all part of extensive listening.

2. Designing Assessment Tasks: Intensive Listening.

Once you have determined objectives, your next step is to design the including making decisions about how you will elicit performance and how you expect the test-taker to respond. We will look at tasks that range from intensive listening performance, such as minimal phonemic pair recognition, to extensive comprehension of language in communicative contexts. The focus in this lecture is the micro-skills of intensive listening.

A typical form of intensive listening at this level is the assessment of recognition of phonological and morphological elements of language. A classic test task gives a spoken stimulus and asks test-taker to identify the stimulus from two or more choices, as in the following examples:

Phonemic pair, consonants

<i>Test-takers hear:</i>	He's from California.
<i>Test-takers read:</i>	(a) He's from California. (b) She's from California.

Phonemic pair, vowels

<i>Test-takers hear:</i>	Is he living?
<i>Test-takers read:</i>	(a) Is he leaving? (b) Is he living?

3. Designing Assessment Tasks: Responsive Listening.

A question-and-answer format can provide some interactivity in these lower-end listening tasks. The test-taker's response is the appropriate answer to a question.

Appropriate response to a question

<i>Test-takers hear:</i>	How much time did you take to do your homework?
<i>Test-takers read:</i>	(a) In about an hour. (b) About an hour. (c) About \$10. (d) Yes, I did.

The objective of this item is recognition of the *wh*-question *how much* and its appropriate response. Distractors are chosen to represent common learner errors: (a) responding to *hOW much* vs. *hOW much longer*;

(b) confusing *how much* in reference to time vs. the more frequent reference to money;

(c) confusing a *wh*-question with a *yes/no* question.

None of the tasks so far discussed have to be framed in a multiple-choice format. They can be offered in a more open-ended framework in which test-takers write or speak the response. The above item would then look like this:

Open-ended response to a question

<i>Test-takers hear:</i>	How much time did you take to do your homework?
<i>Test-takers write or speak:</i>	_____

If open-ended response formats gain a small amount of authenticity and creativity, they of course suffer some in their practicality, as teachers must then read students' responses and judge their appropriateness, which takes time.

4. Designing Assessment Tasks: Selective Listening.

A third type of listening performance is **selective** listening, in which the test-taker listens to a limited quantify of aural input and must discern within it some specific information. A number of techniques have been used that require selective listening.

Listening Cloze. Listening cloze tasks (sometimes called *cloze dictations* or *partial dictations*) require the test-taker to listen to a story, monologue, or conversation and simultaneously read the written text in which selected words or phrases have been deleted. In a listening cloze task, test-takers see a transcript of the passage that they are listening to and fill in the blanks with the words or phrases that they hear.

Information Transfer. Selective listening can also be assessed through an information transfer technique in which aurally processed information must be transferred to a visual representation, such as labeling a diagram, identifying an element in a picture, completing a form, or showing routes on a map.

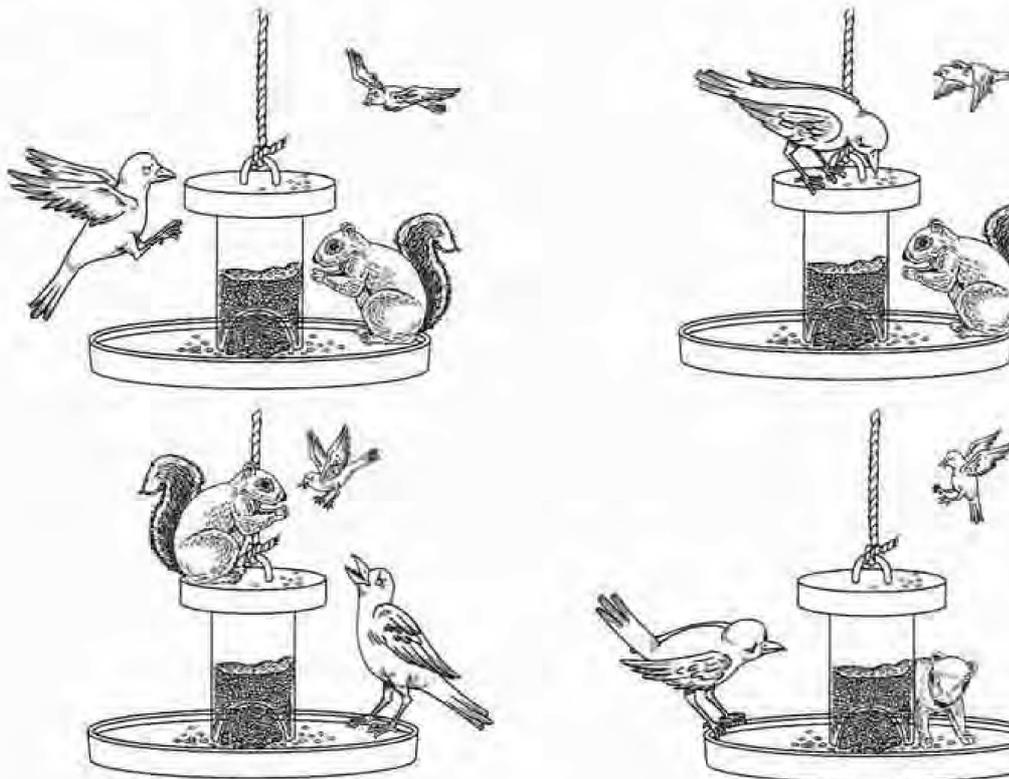
At the lower end of the scale of linguistic complexity, simple picture-cued items are sometimes efficient rubrics for assessing certain selected information. Consider the following item:

Information transfer: multiple-picture-cued selection

Test-takers hear:

Choose the correct picture. In my back yard I have a bird feeder. Yesterday, there two birds and a squirrel fighting for the last few seeds in the bird feeder. The squirrel was on top of the bird feeder while the larger bird sat at the bottom of the feeder screeching at the squirrel. The smaller bird was flying around the squirrel, trying to scare it away.

Test-takers see:



5. Designing Assessment Tasks: Extensive Listening.

Drawing a clear distinction between any two of the categories of listening referred to here is problematic, but perhaps the fuzziest division is between selective extensive listening. We will try to analyze a number of extensive listening comprehension tasks.

Dictation. Dictation is a widely researched genre of assessing listening comprehension. In a dictation, test-takers hear a passage, typically of 50 to 100 words, recited three times: first, at normal speed; then, with long pauses between phrases or natural word groups, during which time test-takers write down what they have just heard; and finally, at normal speed once more so they can check their work and proofread. Here is a sample dictation at the intermediate level of English.

Dictation

First reading (natural speed, no pauses, test-takers listen for gist):

The state of California has many geographical areas. On the western side is the Pacific Ocean with its beaches and sea life. The central part of the state is a large fertile valley. The southeast has a hot desert, and north and west have beautiful mountains and forests. Southern California is a large urban area populated by millions of people.

Second reading (slowed speed, pause at each // break, test-takers write):

The state of California // has many geographical areas. // On the western side // is the Pacific Ocean // with its beaches and sea life. // The central part of the state // is a large fertile valley. // The southeast has a hot desert, // and north and west // have beautiful mountains and forests. // Southern California // is a large urban area // populated by millions of people.

Third reading (natural speed, test-takers check their work).

Scoring is following. Depending on your context and purpose of a dictation, you will need to decide on scoring criteria for several possible kinds of errors:

- spelling error only, but the word appears to have been heard correctly;
- spelling and/or obvious misrepresentation of a word, illegible word;
- grammatical error;
- skipped word or phrase;
- permutation of words;
- additional words not in the original;
- replacement of a word with an appropriate synonym.

Communicative Stimulus-Response Tasks. Another—and more authentic—example of extensive listening is found in a genre of assessment task in which the test-taker is presented with a stimulus monologue or conversation and then is asked to respond to a set of comprehension questions. Such tasks are commonly used in commercially produced proficiency tests. The monologues, lectures, and brief conversations used in such tasks are sometimes a little contrived and certainly the subsequent multiple-choice questions don't mirror communicative, real-life situations. But with some care and creativity, one can create reasonably authentic stimuli, and in some rare cases the response mode actually approaches complete authenticity. Here is a typical example of such a task.

Dialogue and multiple-choice comprehension items

Test-takers hear:

Directions: Now you will hear a conversation between Lynn and her doctor. You will hear the conversation two times. After you hear the conversation the second time choose the correct answer for questions 11–15 below. Mark your answers on the answer sheet provided.

- Doctor: Good morning, Lynn. What's the problem?
Lynn: Well, you see, I have a terrible headache, my nose is running, and really dizzy.
Doctor: Okay. Anything else?
Lynn: I've been coughing, I think I have a fever, and my stomach aches.
Doctor: I see. When did this start?
Lynn: Well, let's see, I went to the lake last weekend, and after I returned home I started sneezing.
Doctor: Hmm. You must have the flu. You should get lots of rest, drink hot beverages, and stay warm. Do you follow me?
Lynn: Well, uh, yeah, but . . . shouldn't I take some medicine?
Doctor: Sleep and rest are as good as medicine when you have the flu.
Lynn: Okay, thanks, Dr. Brown.

Test-takers read:

- 11.** What is Lynn's problem?
(A) She feels horrible.
(B) She ran too fast at the lake.
(C) She's been drinking too many hot beverages.
- 12.** When did Lynn's problem start?
(A) When she saw her doctor.
(B) Before she went to the lake.
(C) After she came home from the lake.
- 13.** The doctor said that Lynn _____.
(A) flew to the lake last weekend
(B) must not get the flu
(C) probably has the flu

14. The doctor told Lynn _____.
- (A) to rest
 - (B) to follow him
 - (C) to take some medicine
15. According to Dr. Brown, sleep and rest are _____ medicine when you have the flu.
- (A) more effective than
 - (B) as effective as
 - (C) less effective than

Authentic Listening Tasks. Ideally, the language assessment field would have a stockpile of listening test types that are cognitively demanding, communicative, and authentic, not to mention integrated with speaking. Here are some possibilities.

1. *Note-taking.* In the academic world, classroom lectures by professors are common features of a non-native English-user's experience. One form of a midterm examination at the American Language Institute at San Francisco State University (Kahn, 2002) uses a 15-minute lecture as a stimulus. One among several response formats includes note-taking by the test-takers. These notes are evaluated by the teacher on a 30-point system, as follows:

Scoring system for lecture notes

0–15 points

Visual representation: Are your notes clear and easy to read? Can you easily find and retrieve information from them? Do you use the space on the page to visually represent ideas? Do you use indentation, headers, numbers, etc.?

0–10 points

Accuracy: Do you accurately indicate main ideas from lectures? Do you note important details and supporting information and examples? Do you leave out unimportant information and tangents?

0–5 points

Symbols and abbreviations: Do you use symbols and abbreviations as much as possible to save time? Do you avoid writing out whole words, and do you not write down every single word the lecturer says?

2. *Editing.* Another authentic task provides both a written and a spoken stimulus, and requires the test-taker to listen for discrepancies. Scoring achieves relatively high reliability as there are usually a small number of specific differences that must be identified. Here is the way the task proceeds.

Editing a written version of an aural stimulus

Test-takers read: the written stimulus material (a news report, an email from a friend, notes from a lecture, or an editorial in a newspaper).

Test-takers hear: a spoken version of the stimulus that deviates, in a finite number of facts or opinions, from the original written form.

Test-takers mark: the written stimulus by circling any words, phrases, facts or opinions that show a discrepancy between the two versions.

3. Interpretive tasks. One of the intensive listening tasks described above was paraphrasing a story or conversation. An interpretive task extends the stimulus material to a longer stretch of discourse and forces the test-taker to infer a response. Potential stimuli include:

- 1 song lyrics;
- 2 recited poetry;
- 3 radio/television news reports;
- 4 an oral account of an experience;

Test-takers are then directed to interpret the stimulus by answering a few questions (in open-ended form). Questions might be:

- 65535 "Why was the singer feeling sad?"
- 65536 "What events might have led up to the reciting of this poem?"
- 65537 "What do you think the political activists might do next, and why?"
- 65538 "What do you think the storyteller felt about the mysterious disappearance of her necklace?"

4. Retelling. In a related task, test-takers listen to a story or news event and simply retell it, or summarize it, either orally (on an audiotape) or in writing. Test-takers must identify the gist, main idea, purpose, supporting point and/or conclusion to show full comprehension. Scoring is partially predetermined by specifying a minimum number of elements that must appear in the retelling. Validity, cognitive processing, communicative ability and authenticity are all well incorporated into the task.

SUGGESTED LITERATURE

1. Buck, Gary. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
2. Richards, Jack C. (1983) Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17, 219-239.
3. Mendelsohn, David J. (1998). Teaching listening. *Annual Review of Applied Linguistics*, 18, 81-101.

LECTURE 3 (CONTINUATION). ASSESSING SPEAKING

Plan:

Basic Types of Speaking

Designing Assessment Tasks: Imitative Speaking.

Designing Assessment Tasks: Intensive Speaking.

Designing Assessment Tasks: Responsive Speaking.

Designing Assessment Tasks: Interactive Speaking.

Designing Assessment Tasks: Extensive Speaking.

Glossary: *imitative speaking, intensive speaking, responsive speaking, interactive speaking, extensive speaking*

1. Basic Types of Speaking

From a pragmatic view of language performance, listening and speaking are almost always closely interrelated. While speaking is a productive skill that can be directly and empirically observed, those observations are invariably colored by the accuracy and effectiveness of a test-taker's listening skill, which necessarily compromises the reliability and validity of an oral production test.

During today's lecture we will cite four categories of speaking assessment tasks. They are:

Imitative Speaking. At one end of a continuum of types of speaking performance is the ability to simply parrot back (imitate) a word or phrase or possibly a sentence. 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.

Intensive Speaking. A second type of speaking frequently employed in assessment contexts is the production of short stretches of oral language designed to demonstrate competence in a narrow band of grammatical, phrasal, lexical, or phonological relationships (such as prosodic elements—intonation, stress, rhythm, juncture). Examples of intensive assessment tasks include directed response tasks, reading aloud, sentence and dialogue completion; limited picture-cued tasks including simple sequences; and translation up to the simple sentence level.

Responsive Speaking. Responsive assessment tasks include interaction and test comprehension but at the somewhat limited level of very short conversations, standard greetings and small talk, simple requests and comments, and the like.

Interactive Speaking. Interaction can take the two forms of transactional language, which has the purpose of exchanging specific information or interpersonal exchanges, which have the purpose of maintaining social relationships. In interpersonal exchanges, oral production can become pragmatically complex with the need to speak in a casual register and use colloquial language, ellipsis, slang, humor, and other sociolinguistic conventions.

Extensive Speaking. Extensive oral production tasks include speeches, oral presentations, and story-telling, during which the opportunity for oral interaction from listeners is either highly limited (perhaps to nonverbal responses) or ruled out altogether.

2. Designing Assessment Tasks: Imitative Speaking.

You may be surprised to see the inclusion of simple phonological imitation in a consideration of assessment of oral production. After all, endless repeating of words, phrases, and sentences was the province of the long-since-discarded Audio-Lingual Method, and in an era of communicative language teaching, many believe that non-meaningful imitation of sounds is fruitless. Such opinions have faded in recent years as we discovered that an overemphasis on fluency can sometimes lead to the decline of accuracy in speech. And so we have been paying more attention to pronunciation, especially suprasegmentals, in an attempt to help learners be more comprehensible.

Word repetition task

<i>Test-takers hear:</i>	<i>Repeat after me:</i>	
	beat [pause] bit [pause]	
	bat [pause] vat [pause]	etc.
	I bought a boat yesterday.	
	The glow of the candle is growing.	etc.
	When did they go on vacation?	
	Do you like coffee?	etc.
<i>Test-takers repeat the stimulus.</i>		

Scoring scale for repetition tasks

2	acceptable pronunciation
1	comprehensible, partially correct pronunciation
0	silence, seriously incorrect pronunciation

3. Designing Assessment Tasks: Intensive Speaking.

At the intensive level, test-takers are prompted to produce short stretches of discourse (no more than a sentence) through which they demonstrate linguistic ability at a specified level of language. Many tasks are “cued” tasks in that they lead the test-taker into a narrow band of possibilities. Intensive tasks may also be described as **limited response tasks** (Madsen, 1983), or **mechanical tasks** (Underbill, 1987) or what classroom pedagogy would label as **controlled responses**.

Directed Response Tasks. In this type of task, the test administrator elicits a particular grammatical form or a transformation of a sentence. Such tasks are clearly mechanical and not communicative, but they do require minimal processing of meaning in order to produce the correct grammatical output.

Directed response

Test-takers hear:

- Tell me he went home.
- Tell me that you like rock music.
- Tell me that you aren't interested in tennis.
- Tell him to come to my office at noon.
- Remind him what time it is.

Read-Aloud Tasks. Intensive reading-aloud tasks include reading beyond the sentence level up to a paragraph or two. This technique is easily administered by selecting a passage that incorporates test specs and by recording the test-taker's output; the scoring is relatively easy because all of the test-taker's oral production is controlled.

Read-aloud stimulus, paragraph length

Despite the decrease in size—and, some would say, quality—of our cultural world, still remain strong differences between the usual British and American writing style question is, how do you get your message across? English prose conveys its most ideas as if they were timeless truths, while American writing exaggerates; if you be half of what is said, that's enough. The former uses understatement; the latter, overstatement. There are also disadvantages to each characteristic approach. Readers who are used to being screamed at may not listen when someone chooses to whisper politely. At the same time, the individual who is used to a quiet manner may reject a series of loud imperatives.

The scoring scale for this passage provided a four-point scale for pronunciation, for fluency, as shown in the box below.

Test of Spoken English scoring scale (1987, p. 10)

Pronunciation:

Points:

- 0.0–0.4 Frequent phonemic errors and foreign stress and intonation patterns that cause the speaker to be unintelligible.
- 0.5–1.4 Frequent phonemic errors and foreign stress and intonation patterns that cause the speaker to be occasionally unintelligible.
- 1.5–2.4 Some consistent phonemic errors and foreign stress and intonation patterns, but the speaker is intelligible.
- 2.5–3.0 Occasional non-native pronunciation errors, but the speaker is always intelligible.

Fluency:

Points:

0.0–0.4	Speech is so halting and fragmentary or has such a non-native flow that intelligibility is virtually impossible.
0.5–1.4	Numerous non-native pauses and/or a non-native flow that interferes with intelligibility.
1.5–2.4	Some non-native pauses but with a more nearly native flow so that the pauses do not interfere with intelligibility.
2.5–3.0	Speech is smooth and effortless, closely approximating that of a native speaker.

Sentence/Dialogue Completion Tasks and Oral Questionnaires. Another technique for targeting intensive aspects of language requires test-takers to read dialogue in which one speaker's lines have been omitted. Test-takers are first given time to read through the dialogue to get its gist and to think about appropriate lines to fill in. Then as the tape, teacher, or test administrator produces one part orally, the test-taker responds. Here's an example.

Dialogue completion task

Test-takers read (and then hear):

In a department store:

Salesperson: May I help you?

Customer: _____.

Salesperson: Okay, what size do you wear?

Customer: _____.

Salesperson: Hmm. How about this green sweater here?

Customer: _____.

Salesperson: Oh. Well, if you don't like green, what color would you like?

Customer: _____.

Salesperson: How about this one?

Customer: _____.

Salesperson: Great!

Customer: _____.

Salesperson: It's on sale today for \$39.95.

Customer: _____.

Salesperson: Sure, we take Visa, MasterCard, and American Express.

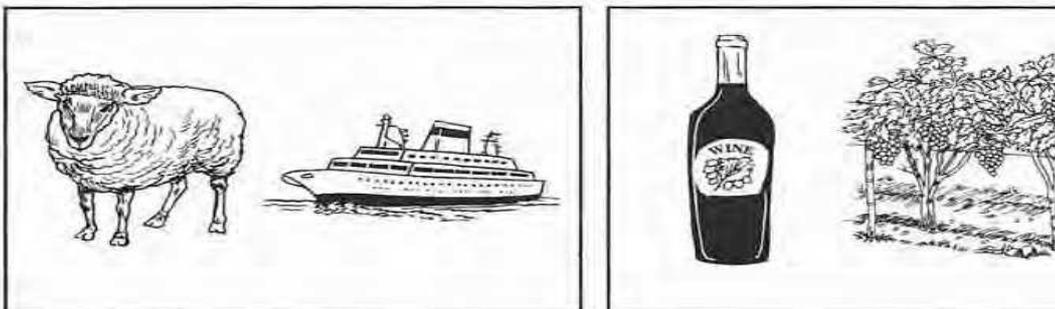
Customer: _____.

Test-takers respond with appropriate lines.

Picture-Cued Tasks. One of the more popular ways to elicit oral language performance at both intensive and extensive levels is a picture-cued stimulus that requires a description from the test-taker. Pictures may be very simple, designed to elicit a word or a phrase; somewhat more elaborate and "busy"; or composed of a series that tells a story or incident. Here is an example of a picture-cued elicitation of the production of a simple minimal pair.

Picture-cued elicitation of minimal pairs

Test-takers see:



Test-takers hear: [test administrator points to each picture in succession]
What's this?

The future tense is elicited with the following picture:

Picture-cued elicitation of future tense (Brown & Sahni, 1994, p. 145)

Test-takers see:



Test-takers hear: This family is at an airport going on their vacation.

1. [point to the picture in general] Where are they going for their vacation?
2. [point to the father] What will he do in Hawaii?
3. [point to the mother] What will she do there?
4. [point to the girl] What is she going to do there?
5. [point to the boy] What is he going to do in Hawaii?

Scoring responses on picture-cued intensive speaking tasks varies, depending on the expected performance criteria. The first task above that asked just for one-word or simple-sentence responses can be evaluated simply as "correct" or "incorrect." The second rubric may apply as well, with these modifications:

Scoring scale for intensive tasks

2	comprehensible; acceptable target form
1	comprehensible; partially correct target form
0	silence, or seriously incorrect target form

4. Designing Assessment Tasks: Responsive Speaking.

Assessment of responsive tasks involves brief interactions with an interlocutor, differing from intensive tasks in the increased creativity given to the test-taker and from interactive tasks by the somewhat limited length of utterances.

Question and Answer. Question-and-answer tasks can consist of one or two questions from an interviewer, or they can make up a portion of a whole battery of questions and prompts in an oral interview. Responsive questions may take the following forms:

Questions eliciting open-ended responses

Test-takers hear:

1. What do you think about the weather today?
2. What do you like about the English language?
3. Why did you choose your academic major?
4. What kind of strategies have you used to help you learn English?
5. a. Have you ever been to the United States before?
b. What other countries have you visited?
c. Why did you go there? What did you like best about it?
d. If you could go back, what would you like to do or see?
e. What country would you like to visit next, and why?

Test-takers respond with a few sentences at most.

Giving Instructions and Directions. We are all called on in our daily routines to read instructions on how to operate an appliance, how to put a bookshelf together, or how to create a delicious clam chowder. The technique is simple: the administrator poses the problem, and the test-taker responds. Scoring is based primarily on comprehensibility and secondarily on other specified grammatical or discourse categories. Here are some possibilities.

Eliciting instructions or directions

Test-takers hear:

- Describe how to make a typical dish from your country.
- What's a good recipe for making _____?
- How do you access email on a PC computer?
- How would I make a typical costume for a _____ celebration in your country?
- How do you program telephone numbers into a cell (mobile) phone?
- How do I get from _____ to _____ in your city?

Test-takers respond with appropriate instructions/directions.

Paraphrasing. Another type of assessment task that can be categorized as responsive asks the test-taker to read or hear a limited number of sentences (perhaps two to five) and produce a paraphrase of the sentence. For example:

Paraphrasing a story

Test-takers hear: Paraphrase the following little story in your own words.

My weekend in the mountains was fabulous. The first day we backpacked into the mountains and climbed about 2,000 feet. The hike was strenuous but exhilarating. At sunset we found these beautiful alpine lakes and made camp there. The sunset was amazingly beautiful. The next two days we just kicked back and did little day hikes with some rock climbing, bird watching, swimming, and fishing. The hike out on the next day was really easy—all downhill—and the scenery was incredible.

Test-takers respond with two or three sentences.

5. Designing Assessment Tasks: Interactive Speaking.

The final two categories of oral production assessment (interactive and extensive speaking) include tasks that involve relatively long stretches of interactive discourse (interviews, role plays, discussions, games) and tasks of equally long duration but that involve less interaction (speeches, telling longer stories, and extended explanations and translations). The obvious difference between the two sets of tasks is the degree of interaction with an interlocutor.

Interview. When "oral production assessment" is mentioned, the first thing that comes to mind is an oral interview: a test administrator and a test-taker sit down in a direct face-to-face exchange and proceed through a protocol of questions and directives.

The suggested set of content specifications for an oral interview (below) may serve as sample questions that can be adapted to individual situations.

Oral interview content specifications

Warm-up:

1. Small talk

Level check:

The test-taker . . .

2. answers *wh*-questions.
3. produces a narrative without interruptions.
4. reads a passage aloud.
5. tells how to make something or do something.
6. engages in a brief, controlled, guided role play.

Probe:

The test-taker . . .

7. responds to interviewer's questions about something the test-taker know and is planning to include in an article or paper.
8. talks about his or her own field of study or profession.
9. engages in a longer, more open-ended role play (for example, simulate a difficult or embarrassing circumstance) with the interviewer.
10. gives an impromptu presentation on some aspect of test-taker's field.

Wind-down:

11. Feelings about the interview, information on results, further questions

Here are some possible questions, probes, and comments that fit those specifications.

Sample questions for the four stages of an oral interview

1. Warm-up:

How are you?
What's your name?
What country are you from? What [city, town]?
Let me tell you about this interview.

2. Level check:

How long have you been in this [country, city]?
Tell me about your family.

What is your [academic major, professional interest, job]?
How long have you been working at your [degree, job]?
Describe your home [city, town] to me.
How do you like your home [city, town]?
What are your hobbies or interests? (What do you do in your spare time?)
Why do you like your [hobby, interest]?
Have you traveled to another country besides this one and your home country?
Tell me about that country.
Compare your home [city, town] to another [city, town].
What is your favorite food?
Tell me how to [make, do] something you know well.
What will you be doing ten years from now?
I'd like you to ask me some questions.
Tell me about an exciting or interesting experience you've had.
Read the following paragraph, please. *[test-taker reads aloud]*
Pretend that you are _____ and I am a _____. *[guided role play follows]*

3. Probe:

What are your goals for learning English in this program?
Describe your [academic field, job] to me. What do you like and dislike about it?
What is your opinion of [a recent headline news event]?
Describe someone you greatly respect, and tell me why you respect that person.
If you could redo your education all over again, what would you do differently?
How do eating habits and customs reflect the culture of the people of your country?
If you were [president, prime minister] of your country, what would you like to change about your country?
What career advice would you give to your younger friends?
Imagine you are writing an article on a topic you don't know very much about. Ask me some questions about that topic.
You are in a shop that sells expensive glassware. Accidentally you knock over an expensive vase, and it breaks. What will you say to the store owner? *[Interviewer role-plays the store owner]*

4. Wind-down:

Did you feel okay about this interview?
What are your plans for [the weekend, the rest of today, the future]?
You'll get your results from this interview [tomorrow, next week].
Do you have any questions you want to ask me?
It was interesting to talk with you. Best wishes.

Table 7.2. Oral proficiency scoring categories (Brown, 2001, pp. 406–407)

	Grammar	Vocabulary	Comprehension
I	Errors in grammar are frequent, but speaker can be understood by a native speaker used to dealing with foreigners attempting to speak his language.	Speaking vocabulary inadequate to express anything but the most elementary needs.	Within the scope of limited language experience can understand simple questions and statements delivered with slow repetition, or paraphrase.
II	Can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.	Has speaking vocabulary sufficient to express himself simply with some circumlocutions.	Can get the gist of most conversations of non-specialized subjects (i.e., topics of no specialized knowledge).
III	Control of grammar is good. Able to speak the language with sufficient structural accuracy to participate effectively in most formal and informal conversations on practical, social, and professional topics.	Able to speak the language with sufficient vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Vocabulary is broad enough that he rarely has to grope for a word.	Comprehension is quite complete at a normal rate of speech.
IV	Able to use the language accurately on all levels normally pertinent to professional needs. Errors in grammar are quite rare.	Can understand and participate in any conversation within the range of his experience with a high degree of precision of vocabulary.	Can understand any conversation within the range of his experience.
V	Equivalent to that of an educated native speaker.	Speech on all levels is fully accepted by educated native speakers in all its features including breadth of vocabulary and idioms, colloquialisms, and pertinent cultural references.	Equivalent to that of an educated native speaker.

Fluency	Pronunciation	Task
(No specific fluency description. Refer to other four language areas for implied level of fluency.)	Errors in pronunciation are frequent but can be understood by a native speaker used to dealing with foreigners attempting to speak his language.	Can ask and answer on topics very familiar. Able to satisfy routine needs and minimum requirements. (Should be able to order a simple meal, shelter or lodging, ask for simple directions, make purchases, and tell time.)
Can handle with confidence but not with facility most social situations, including introductions and casual conversations about current events, as well as work, family, and autobiographical information.	Accent is intelligible though often quite faulty.	Able to satisfy routine social demands and work requirements; needs help handling any complex difficulties.
Can discuss particular interests of competence with reasonable ease. Rarely has to grope for words.	Errors never interfere with understanding and rarely disturb the native speaker. Accent may be obviously foreign.	Can participate effectively in most formal and informal conversations on practical, social, and professional topics.
Able to use the language fluently on all levels normally pertinent to professional needs. Can participate in any conversation within the range of this experience with a high degree of fluency.	Errors in pronunciation are quite rare.	Would rarely be taken as a native speaker but can communicate appropriately even in unfamiliar situations. Can handle informal interaction from and into languages.
Has complete fluency in the language such that his speech is fully accepted by educated native speakers.	Equivalent to and fully accepted by educated native speakers.	Speaking proficiency is equivalent to that of an educated native speaker.

Table 7.3. Subcategories of oral proficiency scores

Level	Description
0	Unable to function in the spoken language
0+	Able to satisfy immediate needs using rehearsed utterances
1	Able to satisfy minimum courtesy requirements and maintain very simple face-to-face conversations on familiar topics
1+	Can initiate and maintain predictable face-to-face conversations and satisfy limited social demands
2	Able to satisfy routine social demands and limited work requirements
2+	Able to satisfy most work requirements with language usage that is often, but not always, acceptable and effective
3	Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics
3+	Often able to use the language to satisfy professional needs in a wide range of social and demanding tasks
4	Able to use the language fluently and accurately on all levels normally pertinent to professional needs
4+	Speaking proficiency is regularly superior in all respects, usually equivalent to that of a well-educated, highly articulate native speaker
5	Speaking proficiency is functionally equivalent to that of a highly articulate, well-educated native speaker and reflects the cultural standards of the country whose language is spoken

Role Play. Role playing is a popular pedagogical activity in communicative language-teaching classes. As an assessment device, role play opens some windows of opportunity for test-takers to use discourse that might otherwise be difficult to elicit.

Discussions and Conversations. As formal assessment devices, discussions and conversations with and

among students are difficult to specify and even more difficult to score. But as *informal* techniques to assess learners, they offer a level of authenticity and spontaneity that other assessment techniques may not provide. Discussions may be especially appropriate tasks through which to elicit and observe such abilities as:

- topic nomination, maintenance, and termination;
- attention getting, interrupting, floor holding, control;
- clarifying, questioning, paraphrasing;
- comprehension signals (nodding, "uh-huh", "hmm," etc.);
- negotiating meaning;
- intonation patterns for pragmatic effect;
- kinesics, eye contact, body language;
- politeness, formality, and other sociolinguistic factors.

Games. Among informal assessment devices are a variety of games that directly involve language production. Consider the following types:

Assessment games

1. "Tinkertoy" game: A Tinkertoy (or Lego block) structure is built behind screen. One or two learners are allowed to view the structure. In successive stages of construction, the learners tell "runners" (who can observe the structure) how to re-create the structure. The runners then "builders" behind another screen how to build the structure. The builders may question or confirm as they proceed, but only through the two degrees of separation. Object: re-create the structure as accurately as possible.
2. Crossword puzzles are created in which the names of all members of class are clued by obscure information about them. Each class member must ask questions of others to determine who matches the clues in the puzzle.
3. Information gap grids are created such that class members must conduct mini-interviews of other classmates to fill in boxes, e.g., "born in July," "plays the violin," "has a two-year-old child," etc.
4. City maps are distributed to class members. Predetermined map directions are given to one student who, with a city map in front of him or her, describes the route to a partner, who must then trace the route and give the correct final destination.

Clearly, such tasks have wandered away from the traditional notion of an oral production test and may even be well beyond *assessments*. As assessments, the key is to specify a set of criteria and a reasonably practical and reliable scoring method. The benefit of such an informal assessment may not be as much in a summative evaluation as in its formative nature, with washback for the students.

6. Designing Assessment Tasks: Extensive Speaking.

Extensive speaking tasks involve complex, relatively lengthy stretches of discourse. They are frequently variations on monologues, usually with minimal verbal interaction.

Oral Presentations. In the academic and professional arenas, it would not be uncommon to be called on to present a report, a paper, a marketing plan, a sales idea, a design of a new product, or a method. A summary of oral assessment techniques would therefore be incomplete without some consideration of extensive speaking tasks. For oral presentations, a checklist or grid is a common means of scoring or evaluation.

Oral presentation checklist

Evaluation of oral presentation

Assign a number to each box according to your assessment of the various aspects of the speaker's presentation.

3	Excellent
2	Good
1	Fair
0	Poor

Content:

- The purpose or objective of the presentation was accomplished.
- The introduction was lively and got my attention.
- The main idea or point was clearly stated toward the beginning.
- The supporting points were
 - clearly expressed
 - supported well by facts, argument
- The conclusion restated the main idea or purpose.

Delivery:

- The speaker used gestures and body language well.
- The speaker maintained eye contact with the audience.
- The speaker's language was natural and fluent.
- The speaker's volume of speech was appropriate.
- The speaker's rate of speech was appropriate.
- The speaker's pronunciation was clear and comprehensible.
- The speaker's grammar was correct and didn't prevent understanding.
- The speaker used visual aids, handouts, etc., effectively.
- The speaker showed enthusiasm and interest.
- [If appropriate] The speaker responded to audience questions well.

Picture-Cued Story-Telling. One of the most common techniques for eliciting oral production is through visual pictures, photographs, diagrams, and charts. We have already looked at this elicitation device for intensive tasks, but at this level we consider a picture or a series of pictures as a stimulus for a longer story or description. Consider the following set of pictures:

Picture-cued story-telling task (Brown, 1999, p. 29)

Test-takers see the following six-picture sequence:

Test-takers hear or read: Tell the story that these pictures describe.
Test-takers use the pictures as a sequence of cues to tell a story.

Retelling a Story, News Event. In this type of task, test-takers hear or read a story or news event that they are asked to retell. This differs from the paraphrasing task discussed above in that it is a longer stretch of discourse and a different genre. The objectives in assigning such a task vary from listening comprehension of the original to production of a number of oral discourse features (communicating sequences and relationships of events, stress and emphasis patterns, "expression" in the case of a dramatic story), fluency, and interaction with the hearer. Scoring should of course meet the intended criteria.

Translation (of Extended Prose). Translation of words, phrases, or short sentences was mentioned under the category of intensive speaking. Here, longer texts are presented for the test-taker to read in the native language and then translate into English. Those texts could come in many forms: dialogue, directions for assembly of a product, a synopsis of

a story or play or movie, directions on how to find something on a map, and other genres. The advantage of translation is in the control of the content vocabulary, and, to some extent, the grammatical and discourse features. Criteria for scoring should therefore take into account not only the purpose in stimulating a translation but the possibility of errors that are unrelated to oral productive ability.

SUGGESTED LITERATURE

1. Underbill, Nic. (1987) *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
2. Brown, J.D. (1998). *New ways of classroom assessment*. Alexandria, VA: Teachers of English to Speakers of Other Languages.
3. Celce-Murcia, Marianne, Brinton, Donna, and Goodwin, Janet. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

LECTURE 4. ASSESSING READING

Plan:

Types and Genres of Reading

Designing Assessment Tasks: Perceptive Reading.

Designing Assessment Tasks: Selective Reading.

Designing Assessment Tasks: Interactive Reading.

Designing Assessment Tasks: Extensive Reading.

Glossary: *perceptive reading, selective reading, interactive reading, extensive reading.*

1. Types (Genres) of Reading

In foreign language learning, reading is a skill that teachers simply expect learners to acquire. The assessment of reading ability does not end with the measurement of comprehension. Strategic pathways to full understanding are often important factors to include in assessing learners, especially in the case of most classroom assessments that are formative in nature.

Each type or **genre** of written text has its own set of governing rules and conventions. Consider the following abridged list of common genres, which ultimately form part of the specifications for assessments of reading ability.

Genres of reading

1. Academic reading

- general interest articles (in magazines, newspapers, etc.)
- technical reports (e.g., lab reports), professional journal articles
- reference material (dictionaries, etc.)
- textbooks, theses
- essays, papers
- test directions
- editorials and opinion writing

2. Job-related reading

- messages (e.g., phone messages)
- letters/emails
- memos (e.g., interoffice)
- reports (e.g., job evaluations, project reports)
- schedules, labels, signs, announcements
- forms, applications, questionnaires
- financial documents (bills, invoices, etc.)
- directories (telephone, office, etc.)
- manuals, directions

3. Personal reading

newspapers and magazines
letters, emails, greeting cards, invitations
messages, notes, lists
schedules (train, bus, plane, etc.)
recipes, menus, maps, calendars
advertisements (commercials, want ads)
novels, short stories, jokes, drama, poetry
financial documents (e.g., checks, tax forms, loan applications)
forms, questionnaires, medical reports, immigration documents
comic strips, cartoons

In the previous lectures we saw that both listening and speaking could be subdivided into at least five different types of listening and speaking performance. In the case of reading, variety of performance is derived more from the multiplicity of types of texts (the genres listed above) than from the variety of overt types of performance. Nevertheless, for considering assessment procedures, several types of reading performance are typically identified, and these will serve as organizers of various assessment tasks.

Perceptive. Perceptive reading tasks involve attending to the *components of* larger stretches of discourse, letters, words, punctuation, and other graphemic symbols.

Selective. This category is an artifact of assessment formats. In order to ascertain one's reading recognition of lexical, grammatical, or discourse features of language within a very short stretch of language, certain typical tasks are used: picture-cued tasks, matching, true/false, multiple-choice, etc.

Interactive. The focus of an interactive task is to identify relevant features (lexical, symbolic, grammatical, and discourse) within texts of moderately short length with the objective of retaining the information that is processed.

Extensive. Extensive reading applies to texts of more than a page, up to and including professional articles, essays, technical reports, short stories, and books.) The purposes of assessment usually are to tap into a learner's global understanding of a text, as opposed to asking test-takers to "zoom in" (увеличить масштаб) on small details.

2. Designing Assessment Tasks: Perceptive Reading.

Assessment of basic reading skills may be carried out in a number of different ways.

Reading Aloud. The test-taker sees separate letters, words, and short sentences and reads aloud, one by one, in the presence of an administrator. Since the assessment *reading* comprehension, any recognizable oral approximation of the response is considered correct.

Written Response. The same stimuli are presented, and the test-taker's task is to reproduce the probe in writing. Because of the transfer across different skills here, evaluation of the test-taker's response must be carefully treated.

Multiple-Choice. Multiple-choice responses are not only a matter of choosing one of four or five possible answers. Other formats, some of which are especially useful at the low levels of reading, include same/different, circle the answer, true/false, choose the letter, and matching. Here are some possibilities.

Minimal pair distinction

Test-takers read:* Circle "S" for same or "D" for different.

1.	led	let	S	D
2.	bit	bit	S	D
3.	seat	set	S	D
4.	too	to	S	D

**In the case of very low level learners, the teacher/administrator reads directions.*

Grapheme recognition task

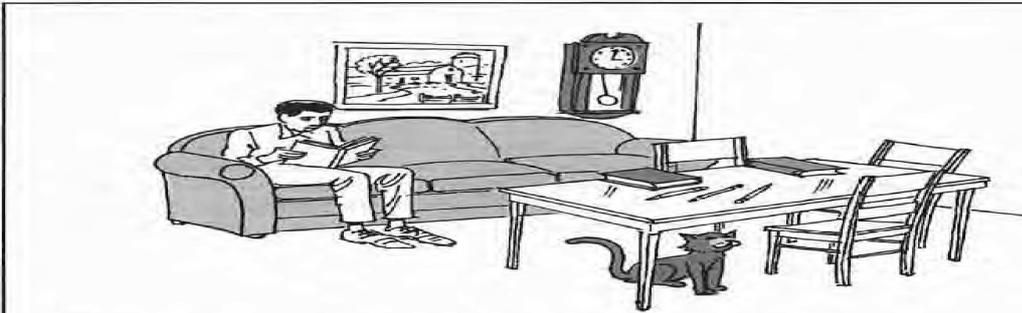
*Test-takers read:** Circle the "odd" item, the one that doesn't "belong."

1. piece peace piece
2. book book boot

**In the case of very low level learners, the teacher/administrator reads directions.*

Picture-Cued Items. Test-takers are shown a picture, such as the one on the next page, along with a written text and are given one of a number of possible tasks to perform.

Picture-cued word identification (Brown & Sahni, 1994, p. 124)



Test-takers hear: Point to the word that you read here.

cat	clock	chai
------------	--------------	-------------

Picture-cued sentence identification

Test-takers hear: Point to the part of the picture that you read about here.

Test-takers see the picture and read each sentence written on a separate card

The man is reading a book.

The cat is under the table.

Picture-cued true/false sentence identification

Test-takers read:

- | | | |
|-------------------------------------|---|---|
| 1. The pencils are under the table. | T | F |
| 2. The cat is on the table. | T | F |
| 3. The picture is over the couch. | T | F |

3. Designing Assessment Tasks: Selective Reading.

Here are some of the possible tasks you can use to assess lexical and grammatical aspects of *reading* ability.

Multiple-Choice (for Form-Focused Criteria). By far the most popular method of testing a reading knowledge of vocabulary and grammar is the multiple-choice format, mainly for reasons of practicality: it is easy to administer and can be scored quickly. The most straightforward multiple-choice items may have little context, but might serve as a vocabulary or grammar check.

Multiple-choice vocabulary/grammar tasks

1. He's not married. He's _____.
A. young
B. single
C. first
D. a husband
2. If there's no doorbell, please _____ on the door.
A. kneel
B. type
C. knock
D. shout
3. The mouse is _____ the bed.
A. under
B. around
C. between
4. The bank robbery occurred _____ I was in the restroom.
A. that
B. during
C. while
D. which

Multiple-choice cloze vocabulary/grammar task

I've lived in the United States (21) _____ three years. I (22) _____ live in C
Rica. I (23) _____ speak any English. I used to (24) _____ homesick, but n
enjoy (25) _____ here. I have never (26) _____ back home (27) _____ I
to the United States, but I might (28) _____ to visit my family soon.

- | | |
|-----------------|-------------|
| 21. A. since | 25. A. live |
| B. for | B. to live |
| C. during | C. living |
| 22. A. used to | 26. A. be |
| B. use to | B. been |
| C. was | C. was |
| 23. A. couldn't | 27. A. when |
| B. could | B. while |
| C. can | C. since |
| 24. A. been | 28. A. go |
| B. be | B. will go |
| C. being | C. going |

Matching Tasks. At this selective level of reading, the test-taker's task is simply to respond correctly, which makes matching an appropriate format. The most frequently appearing criterion in matching procedures is vocabulary. Following is a typical format:

Vocabulary matching task

Write in the letter of the definition on the right that matches the word on the

- | | |
|-----------------------|----------------------------|
| _____ 1. exhausted | a. unhappy |
| _____ 2. disappointed | b. understanding of others |
| _____ 3. enthusiastic | c. tired |
| _____ 4. empathetic | d. excited |

Selected response fill-in vocabulary task

1. At the end of the long race, the runners were totally _____.
2. My parents were _____ with my bad performance on the final exam.
3. Everyone in the office was _____ about the new salary raises.
4. The _____ listening of the counselor made Christina feel well understood.

Choose from among the following:

- disappointed
- empathetic
- exhausted
- enthusiastic

Matching tasks have the advantage of offering an alternative to traditional multiple-choice or fill-in-the-blank formats and are sometimes easier to construct than multiple-choice items, as long as the test designer has chosen the matches carefully.

Editing Tasks. Editing for grammatical or rhetorical errors is a widely used test method for assessing linguistic competence in reading. The TOEFL and many other tests employ this technique with the argument that it not only focuses on grammar but also introduces a simulation of the authentic task of editing, or discerning errors in written passages.

Multiple-choice grammar editing task (Phillips, 2001, p. 219)

Test-takers read: Choose the letter of the underlined word that is not correct.

1. The abrasively action of the wind wears away softer layers of rock.
A B C D

2. There are two way of making a gas condense: cooling it or putting it under pressure.
A B C D

3. Researchers have discovered that the application of bright light can sometimes uses to overcome jet lag.
A B C D

Picture-Cued Tasks. In the previous lecture we looked at picture-cued tasks for perceptive recognition of symbols and words. Pictures and photographs may be equally well utilized for examining ability at the selective level. Several types of picture-cued methods are commonly used.

1. Test-takers read a sentence or passage and choose one of four pictures that is being described. The sentence (or sentences) at this level is more complex. A computer-based example follows:

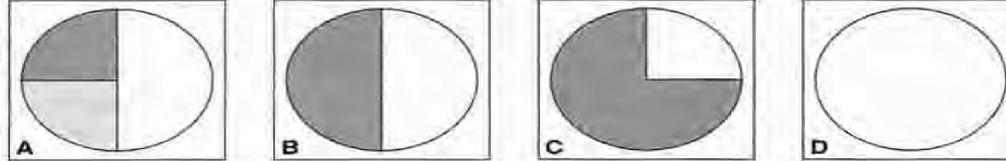
Multiple-choice picture-cued response (Phillips, 2001, p. 276)

Test-takers read a three-paragraph passage, one sentence of which is:

During at least three quarters of the year, the Arctic is frozen.

Click on the chart that shows the relative amount of time each year that water is available to plants in the Arctic.

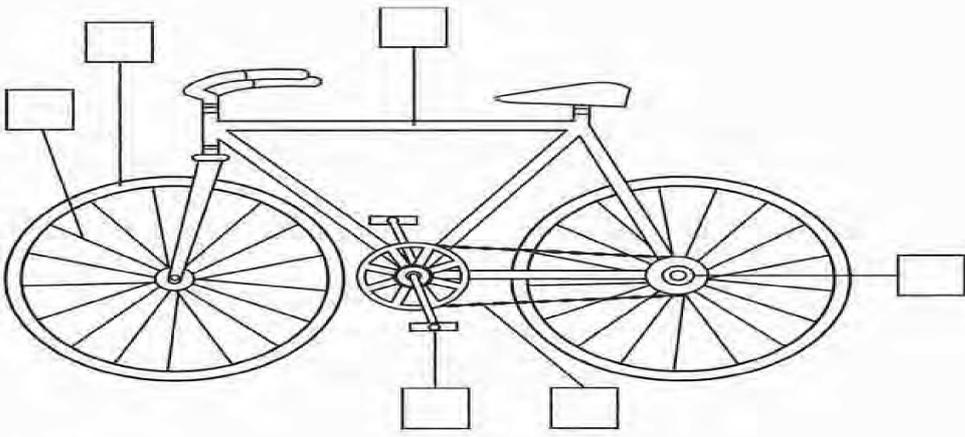
Test-takers see the following four pictures:



2. Test-takers read a series of sentences or definitions, each describing a labeled part of a picture or diagram. Their task is to identify each labeled item. In the following diagram, test-takers do not necessarily know each term, but by reading the definition they are able to make an identification. For example:

Diagram-labeling task

Test-takers see:



Test-takers read:

Label the picture with the number of the corresponding item described below.

1. wire supports extending from the hub of a wheel to its perimeter
2. a long, narrow support pole between the seat and the handlebars
3. a small, geared wheel concentric with the rear wheel
4. a long, linked, flexible metal device that propels the vehicle
5. a small rectangular lever operated by the foot to propel the vehicle
6. a tough but somewhat flexible rubber item that circles each wheel

Gap-Filling Tasks. Many of the multiple-choice tasks described above can be converted into gap-filling or "fill-in-the-blank," items in which the test-taker's response is to write a word or phrase. An extension of simple gap-filling tasks is to create sentence completion items where test-takers read part of a sentence and then complete it by writing a phrase.

Sentence completion tasks

Oscar: Doctor, what should I do if I get sick?
Doctor: It is best to stay home and _____
If you have a fever, _____
You should drink as much _____
The worst thing you can do is _____
You should also _____

4. Designing Assessment Tasks: Interactive Reading.

Tasks at this level, like selective tasks, have a combination of form-focused and meaning-focused objectives but with more emphasis on meaning. Texts are a little longer, from a paragraph to as much as a page or so in the case of ordinary prose. Charts, graphs, and other graphics may be somewhat complex in their format.

Cloze Tasks. One of the most popular types of reading assessment task is the cloze procedure. The word *cloze* was coined by educational psychologists to capture the Gestalt psychological concept of "closure", that is, the ability to fill in gaps in an incomplete image (visual, auditory, or cognitive) and supply (from background schemata) omitted details.

Cloze procedure, rational deletion (prepositions and conjunctions)

The recognition that one's feelings (1) _____ happiness (2) _____ unhappiness can coexist much like love and hate (3) _____ a close relationship may offer valuable clues (4) _____ how to lead a happier life. suggests, (5) _____ example, that changing (6) _____ avoiding things that make you miserable may well make you less miserable (7) _____ probably happier.

Short-Answer Tasks. Multiple-choice items are difficult to construct and validate, and classroom teachers rarely have time in their busy schedules to design such a test. A popular alternative to multiple-choice questions following

reading passages is the age-old short-answer format.

Open-ended reading comprehension questions

1. What do you think the main idea of this passage is?
2. What would you infer from the passage about the future of air travel?
3. In line 6 the word *sensation* is used. From the context, what do you think the word means?
4. What two ideas did the writer suggest for increasing airline business?
5. Why do you think the airlines have recently experienced a decline?

Editing (Longer Texts). The previous part of today's lecture (on selective reading) described editing tasks, but there the discussion was limited to a list of unrelated sentences, each presented with an error to be detected by the test-taker. The same technique has been applied successfully to longer passages of 200 to 300 words. Several advantages are gained in the longer format.

Contextualized grammar editing tasks (Imao, 2001)

- (1) Ever since supermarkets first appeared, they have been take over the world.
A B C D
- (2) Supermarkets have changed people's life styles, yet and at the same time,
A B C
- changes in people's life styles have encouraged the opening of supermarkets. (3
D
- a result this, many small stores have been forced out of business. (4) Moreover,
A B C D B
- small stores will be able to survive this unfavorable situation.
A C D

Scanning. Scanning is a strategy used by all readers to find relevant information in a text. Assessment of scanning is carried out by presenting test-takers with a text and requiring rapid identification of relevant bits of information. Possible stimuli include:

- a one- to two-page news article;
- an essay;
- a chapter in a textbook;
- a technical report;
- a table or chart depicting some research findings;
- a menu;
- an application form.

5. Designing Assessment Tasks: Extensive Reading.

Extensive reading involves somewhat longer texts than we have been dealing with up to this point. Journal articles, technical reports, longer essays, short stories, and books fall into this category. The tasks that are unique to extensive reading: skimming, summarizing, responding to reading, and note-taking.

Skimming Tasks. Skimming is the process of rapid coverage of reading matter to determine its gist or main idea. It is a prediction strategy used to give a reader a sense of the topic and purpose of a text, the organization of the text, the perspective or point of view of the writer, its ease or difficulty. Assessment of skimming strategies is usually straightforward: the test-taker skims a text and answers questions such as the following:

Skimming tasks

- What is the main idea of this text?
What is the author's purpose in writing the text?
What kind of writing is this [newspaper article, manual, novel, etc.]?
What type of writing is this [expository, technical, narrative, etc.]?
How easy or difficult do you think this text will be?
What do you think you will learn from the text?
How useful will the text be for your [profession, academic needs, interests]?

Summarizing and Responding. One of the most common means of assessing extensive reading is to ask

the test-taker to write a summary of the text. The task that is given to students can be very simply worded:

Directions for summarizing

Write a summary of the text. Your summary should be about one paragraph in length (100–150 words) and should include your understanding of the main idea and supporting ideas.

Criteria for assessing a summary (Imao, 2001, p. 184)

1. Expresses accurately the main idea and supporting ideas.
2. Is written in the student's own words; occasional vocabulary from the original text is acceptable.
3. Is logically organized.
4. Displays facility in the use of language to clearly express ideas in the text.

Note-Taking and Outlining. Finally, a reader's comprehension of extensive texts may be assessed through an evaluation of a process of note-taking or outlining. Because of the difficulty of controlling the conditions and time frame for both these techniques, they rest firmly in the category of informal assessment. Their utility is in the strategic training that learners gain in retaining information through marginal notes that highlight key information or organizational outlines that put supporting ideas into a visually manageable framework. A teacher, perhaps in one-on-one conferences with students, can use student notes/outlines as indicators of the presence or absence of effective reading strategies, and thereby point the learners in positive directions.

SUGGESTED LITERATURE

1. Alderson, J. Charles. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
2. Read, John. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
3. Nuttall, Christine. (1996). *Teaching reading skills in a foreign language*. Second Edition. Oxford: Heinemann.

LECTURE 5. ASSESSING WRITING

Plan:

Genres and Types of Written Language.

Designing Assessment Tasks: Imitative Writing.

Designing Assessment Tasks: Intensive (Controlled) Writing.

Designing Assessment Tasks: Responsive Writing and Extensive Writing.

Glossary: *imitative writing, intensive (controlled) writing, responsive writing, extensive writing*

1. Genres and Types of Written Language.

Before looking at specific tasks, we must scrutinize the different genres of written language and types of writing.

Types of writing performance. Four categories of written performance that capture the range of written production are considered here. Each category resembles the categories defined for the other three skills, but these categories, as always, reflect the uniqueness of the skill area.

1. **Imitative.** To produce written language, the learner must attain skills in the fundamental, basic tasks of writing letters, words, punctuation, and very brief sentences. This category includes the ability to spell correctly and to perceive phoneme-grapheme correspondences in the English spelling system.
2. **Intensive (controlled).** Beyond the fundamentals of imitative writing are skills in producing appropriate vocabulary within a context, collocations and idioms and correct grammatical features up to the length of a sentence. Meaning and context are of some importance in determining correctness and appropriateness.

3. **Responsive.** Here, 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 respond 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 or graphs.

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.

2. Designing Assessment Tasks: Imitative Writing.

Many beginning-level English learners, from young children to older adults, need basic training in and assessment of imitative writing: the rudiments of forming letters, words and simple sentences. We examine this level of writing first.

Tasks in Writing Letters, Words, and Punctuation. First, a comment should be made on the increasing use of personal and laptop computers and handheld instruments for creating written symbols. A limited variety of types of tasks are commonly used to assess a person's ability to produce written letters and symbols. A few of the more common types are described here.

1. *Copying.*
2. *Listening cloze selection tasks.*
3. *Picture-cued tasks.*
4. *Form completion tasks.*
5. *Converting numbers and abbreviations to words.*

Spelling Tasks and Detecting Phoneme-Grapheme Correspondences. A number of task types are in popular use to assess the ability to spell words correctly and to process phoneme-grapheme correspondences.

1. **Spelling tests.** The teacher dictates a simple list of words, one word at a time, followed by the word in a sentence repeated again, with a pause for test-takers to write the word. Scoring emphasizes correct spelling.

2. **Picture-cued tasks.** Pictures are displayed with the objective of focusing on familiar words whose spelling may be unpredictable, items are chosen according to the objectives of the assessment, but this format is an opportunity to present some challenging words and word pairs: *boot/book, read/reed, bit/bite*. etc.

3. **Multiple-choice techniques.** Presenting words and phrases in the form of a multiple-choice task risks crossing over into the domain of assessing reading, but if the items have a follow-up writing component, they can serve as formative reinforcement of spelling conventions. They might be more challenging with the addition of homonyms. Here are some examples.

4. **Matching phonetic symbols.** If students have become familiar with the phonetic alphabet, they could be shown phonetic symbols and asked to write the correctly spelled word alphabetically. This works best with letters that do not have one-to-one correspondence with the phonetic symbol (e.g., [ʒ] and a).

3. Designing Assessment Tasks: Intensive (Controlled) Writing.

This next level of writing is what second language teacher training manuals have for decades called controlled writing. It may also be thought of as form-focused writing, grammar writing, or simply guided writing. A good deal of writing at this level is display writing as opposed to real writing: students produce language to display their competence in grammar, vocabulary, or sentence formation, and not necessarily to convey meaning for an authentic purpose.

Dictation and Dicto-Comp. Dictation is the rendition in writing of what one hears aurally, so it could be classified as a type of writing and also dictation can be classified as an intensive form of writing. A form of controlled writing related to dictation is a dicto-comp. Here, a paragraph is read at normal speed, usually two or three times; then the teacher asks students to rewrite the paragraph from the best of their recollection. In one of several variations of the dicto-comp technique, the teacher, after reading the passage, distributes a handout with key words from the paragraph, in sequence, as cues for the students. In either case, the dicto-comp is genuinely classified as an intensive, if not a responsive, writing task. Test-takers must internalize the content of the passage, remember a few phrases and lexical items as key words, then recreate the story in their own words.

Grammatical Transformation Tasks. Language teachers use this technique as an assessment task, ostensibly to measure grammatical competence. Numerous versions of the task are possible:

- Change the tenses in a paragraph.
- Change full forms of verbs to reduced forms (contractions).

- Change statements to *yes/no* or *wh*-questions.
- Change questions into statements.
- Combine two sentences into one using a relative pronoun.
- Change direct speech to indirect speech.
- Change from active to passive voice and etc.

Picture-Cued Tasks. A variety of picture-cued controlled tasks have been used in English classroom around the world. The main advantage in this technique is in detaching the almost ubiquitous reading and writing connection and offering instead a nonverbal means to stimulate written responses. For example:

Vocabulary Assessment Tasks. Most vocabulary study is carried out through reading. The major techniques used to assess vocabulary are defining and using a word in a sentence (multiple choice techniques, matching, picture-cued identification, cloze techniques, guessing the meaning of a word in context).

Read (2000) suggested several types of items for assessment of basic knowledge of the meaning of a word, collocational possibilities, and derived morphological forms. His example centered on the word *interpret*, as follows:

Short-Answer and Sentence Completion Tasks. Such items range from very simple and predictable to somewhat more elaborate responses. Look at range of possibilities.

4. Designing Assessment Tasks: Responsive Writing and Extensive Writing.

Responsive writing creates the opportunity for test-takers to offer an array of possible creative responses within a pedagogical or assessment framework: test-takers are "responding" to a prompt or assignment.

Extensive, or "free" writing takes all the principles and guidelines of responsive writing and puts them into practice in longer texts such as full-length essays, term papers, project reports, and theses and dissertations. In extensive writing, however, the writer has been given, even more freedom to choose: topics, length, style, and perhaps even conventions of formatting are less constrained than in the typical responsive writing exercise.

Let's consider responsive and extensive writing tasks:

Paraphrasing. One of the more difficult concepts for second language learners to grasp is paraphrasing. The initial step in teaching paraphrasing is to ensure that learners understand the importance of paraphrasing: to say something in one's own words, to avoid plagiarizing, to offer some variety in expression. With those possible motivations and purposes in mind, the test designer needs to elicit a paraphrase of a sentence or paragraph.

Scoring of the test-taker's response is a judgment call in which the criterion of conveying the same or similar message is primary, with secondary evaluations of discourse, grammar and vocabulary. Other components of analytic or holistic scales might be considered as criteria for an evaluation.

Guided Question and Answer. Another lower-order task in this type of writing, which has the pedagogical benefit of guiding a learner without dictating the form of the output, is a guided question-and-answer format in which the test administrator poses a series of questions that essentially serve as an outline of the emergent written text. In the writing of a narrative that the teacher has already covered in a class discussion, the following kinds of questions might be posed to stimulate a sequence of sentences.

Paragraph Construction Tasks. The participation of reading performance is inevitable in writing effective paragraphs. To a great extent, writing is the art of emulating what one reads. You read an effective paragraph; you analyze the ingredients of its success; you emulate it. Assessment of paragraph development takes on a number of different forms:

1. Topic sentence writing.
2. Topic development within a paragraph.
3. Development of main and supporting ideas across paragraphs.

At responsive and extensive levels of writing, three major approaches to scoring writing performance are commonly used by test designers: *holistic*, *primary trait* and *analytical*. In the first method, a single score is assigned to an essay, which represents a reader's general overall assessment. Primary trait scoring is a variation of the holistic method in that the achievement of the primary purpose, or trait, of an essay is the only factor rated. Analytical scoring breaks a test-taker's written text down into a number of subcategories (organization, grammar, etc.) and gives a separate rating for each.

SUGGESTED LITERATURE

1. Weigle, Sara Gushing. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
2. Raimes, Ann. (1998). Teaching writing. *Annual Review of Applied Linguistics*, 18. pp. 142-167.