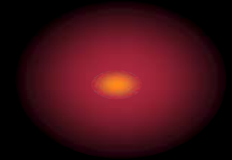
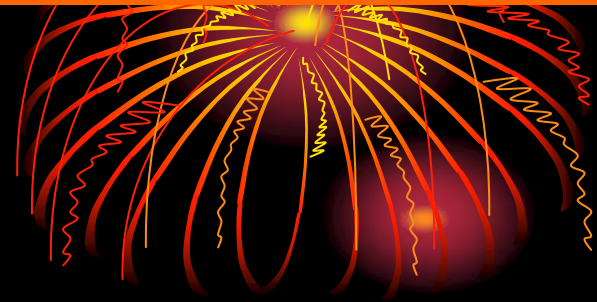


Language Testing and Program Evaluation

Chapter 6 - Part 2
(Brown, 2004)

ASSESSING LISTENING Designing Assessment Tasks

Prof. Dr. Sabri Koç



Designing Assessment Tasks: Intensive Listening

- **Once you have determined objectives, your next step is to design the tasks, including making decisions about how you will elicit performance and how you will expect the test-taker to respond. We will look at tasks that range from intensive listening performance to extensive comprehension of language in communicative contexts. The focus in this section is on the microskills of intensive listening.**



Designing Assessment Tasks: 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-takers to identify the stimulus from two or more choices, as in the following two examples:

Designing Assessment Tasks: Intensive Listening



- Recognizing Phonological and Morphological Elements

a. Phonemic pair, consonants

Test-takers hear: He's from California.

Test-takers read: a. He's from California.

b. She's from California.

b. Phonemic pair, vowels

Test-takers hear:

Is he living?

Test-takers read:

a. Is he leaving?

b. Is he living?



In both cases above, minimal phonemic distinctions are the target. If you are testing recognition of morphology, you can use the same format:

c. Morphological pair, -ed ending

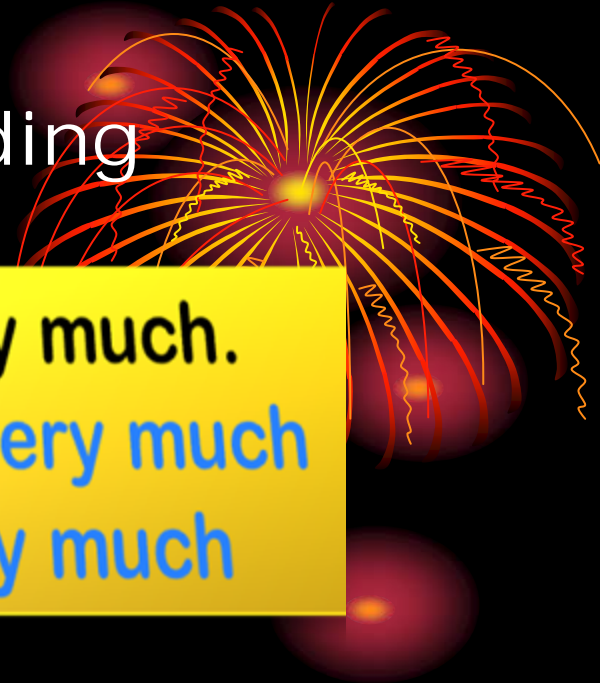
Test-takers hear: I missed you very much.

Test-takers read: a. I missed you very much

b. I miss you very much

Hearing **the past tense morpheme** in this sentence challenges even advanced learners, especially if no context is provided. **Stressed** and **unstressed words** may also be tested with the same rubric.

In the following example, **the reduced form** (contraction) of **can not** is tested:



Stress pattern in *can't*

Test-takers hear: My girl friend can't go to the party?

Test-takers read: a. My girl friend can't go to the party?

b. My girl friend can go to the party?

n' wɪl ði: ɪend ɔn ɡo tə ði: pɑ: ɹ̩: ɹ̩:



- Because they are **decontextualized**, these kinds of tasks leave something to be desired in their authenticity. But they are a step better than items that simply provide **a one-word stimulus**:



One-word stimulus

Test-takers hear: vine

Test-takers read: **a. vine**

b. wine

PARAPHRASE RECOGNITION

Sentence paraphrase

The next step up on the scale of listening comprehension microskills is words, phrases, and sentences, which are frequently assessed by providing a stimulus sentence and asking the test-taker to choose the correct paraphrase from a number of choices.



PARAPHRASE RECOGNITION

Sentence paraphrase

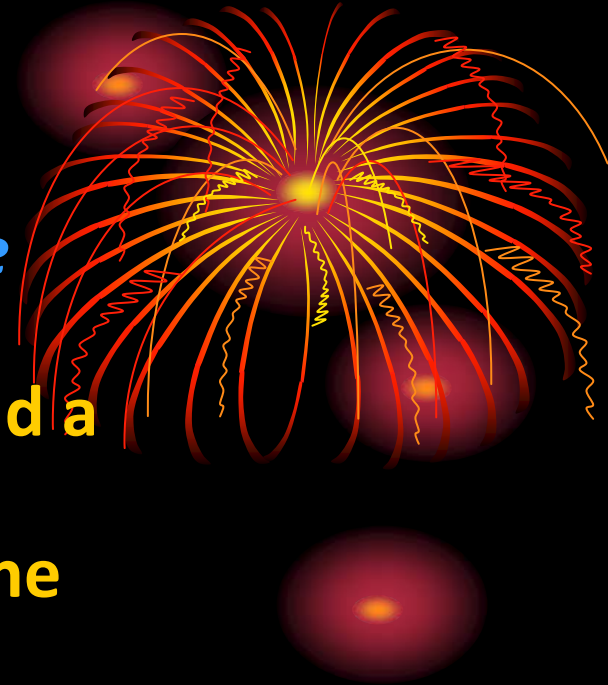


Test-takers hear:

Hellow, my name's Keiko. I come from Japan.

Test-takers read:

- a. Keiko is comfortable in Japan.**
- b. Keiko wants to come to Japan.**
- c. Keiko is Japanese.**
- d. Keiko likes Japan.**



In the above item, the idiomatic *come from* is the phrase being tested. To add a little context, a conversation can be the stimulus task to which test-takers must respond with the correct paraphrase:

Dialog paraphrase

Test-takers hear:

Man : Hi, Maria. My name is George.

Woman : Nice to meet you, George.

Are you American?

Man : No, I'm Canadian.

Test-takers read:

a. George lives in the United States.

b. George is an American.

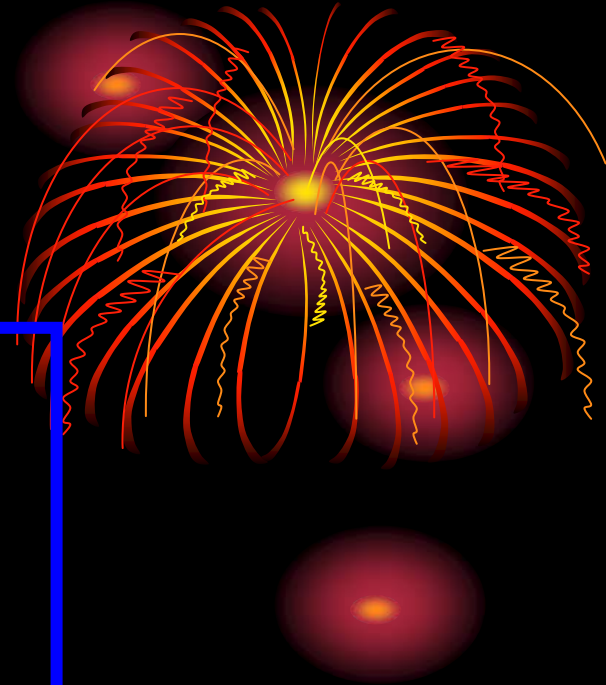
c. George comes from Canada.

d. Maria is Canadian.



Dialog paraphrase

Here, the criterion is
recognition of the adjective
form **used** to indicate country
of origin: **Canadian, American,**
Brazilian, Italian, etc.



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.**

Designing assessment tasks:

Responsive Listening

Appropriate response to a question



Test-takers hear:

How much time did you take to do your homework?

Test-takers read: (a) In about an hour.

(b) About an hour.

(c) About \$10.

(d) Yes, I did.

Designing assessment tasks:

Responsive Listening

Appropriate response to a question



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; (c) confusing how much in reference to time vs. the more frequent reference to money; (d) confusing a wh-question with a yes/no question.

Designing assessment tasks:

Responsive Listening



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

Test-takers hear: How much time did you take to do your homework?

Test-takers write or speak: _____

Designing assessment tasks:

Responsive Listening



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.

DESIGNING ASSESSMENT TASKS:

SELECTIVE LISTENING



In **selective** listening, the test-taker listens to a **limited quantity of aural input** and must discern within it some specific information. A number of techniques have been used that require selective listening.

DESIGNING ASSESSMENT TASKS:

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.

DESIGNING ASSESSMENT TASKS: SELECTIVE LISTENING



Listening Cloze

Cloze procedure is most commonly **associated with reading only**. In its generic form, the test consists of a passage in which every *nth* word (typically every seventh word) is deleted and the test-taker is asked to supply an appropriate word. 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.

DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Listening cloze

Test-takers hear: Ladies and gentlemen, I now have some connecting gate information for those of you making connections to other flights out of San Francisco.

Flight seven-oh-six to Portland will depart from gate seventy-three at nine-thirty P.M.

Flight ten-forty-five to Reno will depart at nine-fifty P.M. from gate seventeen.

Flight four-forty to Monterey will depart at nine-thirty-five P.M. from gate sixty.

And flight sixteen-oh-three to Sacramento will depart from gate nineteen at ten-fifteen P.M.

Test-takers write the missing words or phrases in the blanks.



DESIGNING ASSESSMENT TASKS: SELECTIVE LISTENING

Other listening cloze tasks may focus on a grammatical category such as verb tenses, articles, two-word verbs, prepositions, or transition words/phrases. Notice two important structural differences between listening cloze tasks and standard reading cloze. In a listening cloze, deletions are governed by the objective of the test, not by mathematical deletion of every nth word; and more than one word may be deleted, as in the above example.

Listening cloze tasks should normally use an exact word method of scoring, in which you accept as a correct response only the actual word or phrase that was spoken and consider other appropriate words as incorrect.



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

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:



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: multiple-picture-cued selection

Test-takers hear:

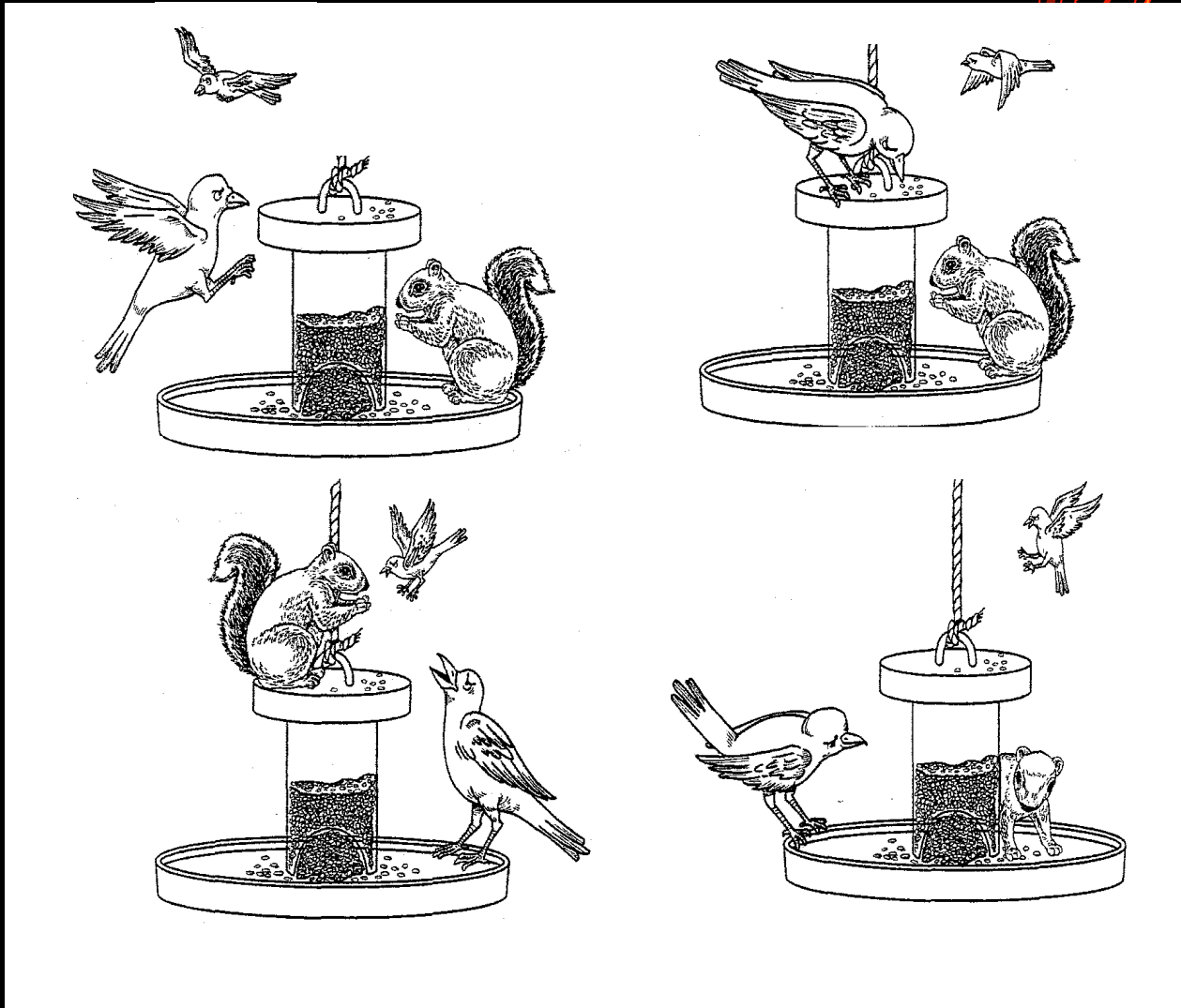
Choose the correct picture. In my back yard I have a bird feeder. Yesterday, there were two birds and a squirrel fighting for the last few seeds in the bird feeder. The squirrel was on top of the bird feeder while the larger bird sat at the bottom of the feeder screeching at the squirrel. The smaller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




DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: multiple-picture-cued selection

Test-takers see:



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**



The preceding example illustrates the need for test-takers to focus on just the relevant information. The objective of this task is to test prepositions and prepositional phrases of location (*at the bottom, on top of, around, along with larger, smaller*), so other words and phrases such as *back yard, yesterday, last few seeds, and scare away* are supplied only as context and need not be tested. The test-takers need to know the difference between **birds** and **squirrels**.

DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: single-picture-cued verbal multiple-choice

Test-takers see: a photograph of a woman in a laboratory setting, with no glasses on, squinting through a microscope with her right eye, and with her left eye closed.

Test-takers hear:

- (a) She's speaking into a microphone.
- (b) She's putting on her glasses.
- (c) She has both eyes open.
- (d) She's using a microscope.



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: single-picture-cued verbal multiple-choice

Information transfer tasks may reflect **greater authenticity** by using **charts, maps, grids, timetables, and other artifacts of daily life**. In the example below, **test-takers hear a student's daily schedule**, and the task is **to fill in the partially completed weekly calendar**.



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: chart-filling

Test -takers hear:

Now you will hear information about Lucy's daily schedule. The information will be given twice. The first time just listen carefully. The second time, there will be a pause after each sentence. Fill in Lucy's blank daily schedule with the correct information. The example has already been filled in.



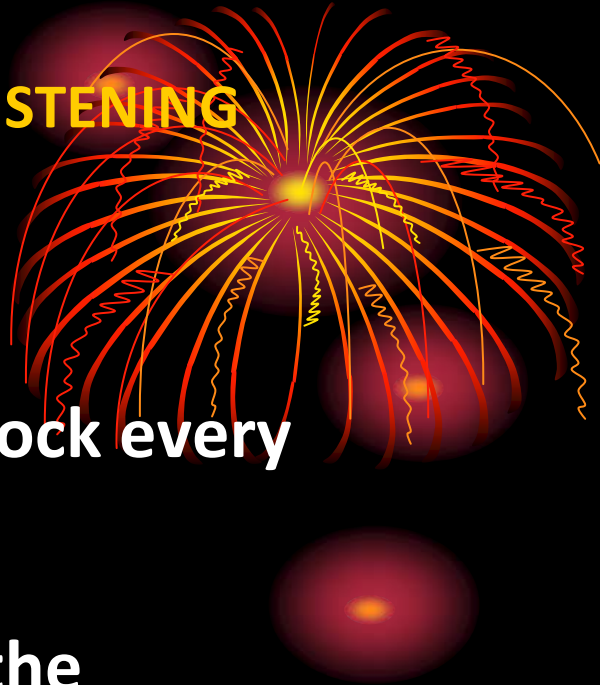
DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: chart-filling

You will hear: Lucy gets up at eight o'clock every morning except on weekends.

You will fill in the schedule to provide the information.

Now listen to the information about Lucy's schedule. Remember, you will first hear all the sentences; then you will hear each sentence separately with time to fill in your chart.



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: chart-filling



Lucy gets up at 8:00 every morning except on weekends. She has English on Monday, Wednesday, and Friday at ten o'clock. She has History on Tuesdays and Thursdays at two o'clock. She takes Chemistry on Monday from two o'clock to six o'clock. She plays tennis on weekends at four o'clock. She eats lunch at twelve o'clock every day except Saturday and Sunday.

DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Information transfer: chart-filling

Lucy gets up at 8:00 every morning except on weekends. She has English on Monday, Wednesday, and Friday at ten o'clock. She has History on Tuesdays and Thursdays at two o'clock. She takes Chemistry on Monday from two o'clock to six o'clock. She plays tennis on weekends at four o'clock. She eats lunch at twelve o'clock every day except Saturday and Sunday.

Test-takers see the following weekly calendar grid:



DESIGNING ASSESSMENT TASKS:

SELECTIVE LISTENING

Information transfer: chart-filling



	Monday	Tuesday	Wednesday	Thursday	Friday	Weekends
8:00	get up	get up	get up	get up	get up	
10:00						
12:00						
2:00						
4:00						
6:00						

DESIGNING ASSESSMENT TASKS:

SELECTIVE LISTENING

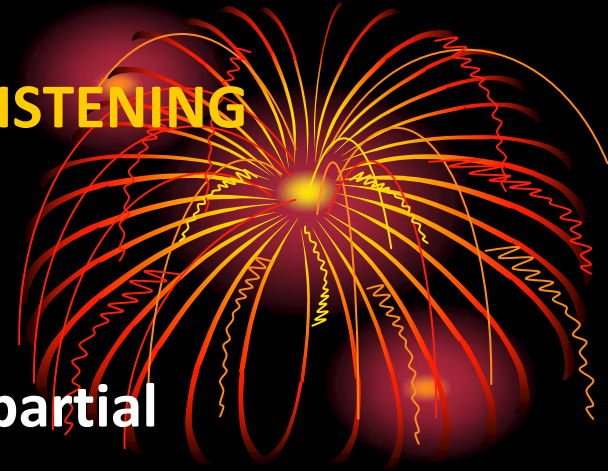
Such chart-filling tasks are good examples of **aural scanning strategies**. A listener must discern from a number of pieces of information which pieces are relevant. In the above example, **virtually all of the stimuli are relevant**, and very few words can be ignored. In other tasks, however, much more information might be presented than is needed, forcing the test-taker to select the correct bits and pieces necessary to complete a task.



DESIGNING ASSESSMENT TASKS: **SELECTIVE LISTENING**

Sentence Repetition

The task of simply **repeating a sentence** or a partial sentence, or sentence repetition, is also used as an **assessment of listening comprehension**. As in a dictation, the test-taker must retain a stretch of language long enough to reproduce it, and then must respond with an oral repetition of that stimulus. Incorrect listening comprehension, whether at the phonemic or discourse level, may be manifested in the correctness of the repetition.



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 and extensive listening. As we gradually move along the continuum from smaller to larger stretches of language, and from micro- to macroskills of listening, the probability of using more extensive listening tasks increases. Some important questions about designing assessments at this level emerge.



DESIGNING ASSESSMENT TASKS: **EXTENSIVE LISTENING**

- 1.** Can listening performance be distinguished from cognitive processing factors such as memory, associations, storage, and recall?
- 2.** As assessment procedures become more communicative, does the task take into account test-takers' ability to use grammatical expectancies, lexical collocations, semantic interpretations, and pragmatic competence?
- 3.** Are test tasks themselves correspondingly content valid and authentic, that is, do they mirror real-world language and context?
- 4.** As assessment tasks become more and more open-ended, they more closely resemble pedagogical tasks, which leads one to ask what the difference is between assessment and teaching tasks. The answer is **scoring**: the former imply specified scoring procedures, while the latter do not.



DESIGNING ASSESSMENT TASKS: **EXTENSIVE LISTENING**



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.

See 03 Brown_Chapter 06_Samples

**THANKS
FOR
LISTENING!**

