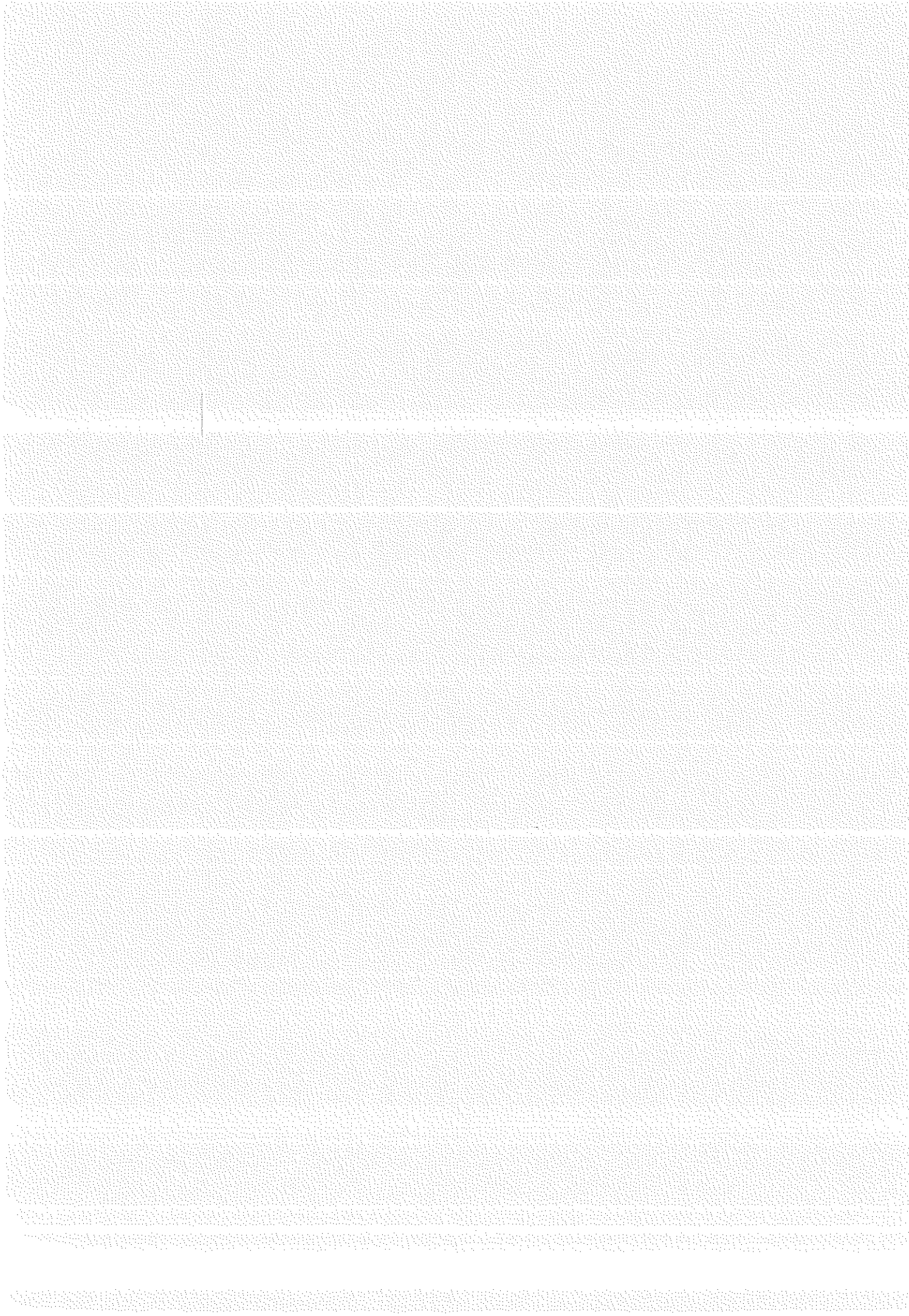


TESTING
SPEAKING SKILL



ASSESSING SPEAKING

From a pragmatic view of language performance, listening and speaking are almost always closely interrelated. While it is possible to isolate some listening performance types (see Chapter 6), it is very difficult to isolate oral production tasks that do not directly involve the interaction of aural comprehension. Only in limited contexts of speaking (monologues, speeches, or telling a story and reading aloud) can we assess oral language without the aural participation of an interlocutor.

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. How do you know for certain that a speaking score is exclusively a measure of oral production without the potentially frequent clarifications of an interlocutor? This interaction of speaking and listening challenges the designer of an oral production test to tease apart, as much as possible, the factors accounted for by aural intake.

Another challenge is the design of elicitation techniques. Because most speaking is the product of creative construction of linguistic strings, the speaker makes choices of lexicon, structure, and discourse. If your goal is to have test-takers demonstrate certain spoken grammatical categories, for example, the stimulus you design must elicit those grammatical categories in ways that prohibit the test-taker from avoiding or paraphrasing and thereby dodging production of the target form.

As tasks become more and more open ended, the freedom of choice given to test-takers creates a challenge in scoring procedures. In receptive performance, the elicitation stimulus can be structured to anticipate predetermined responses and only those responses. In productive performance, the oral or written stimulus must be specific enough to elicit output within an expected range of performance such that scoring or rating procedures apply appropriately. For example, in a picture-series task, the objective of which is to elicit a story in a sequence of events, test-takers could opt for a variety of plausible ways to tell the story, all of which might be equally accurate. How can such disparate responses be evaluated? One solution is to assign not one but several scores for each response, each score representing one of several traits (pronunciation, fluency, vocabulary use, grammar, comprehensibility, etc.).

All of these issues will be addressed in this chapter as we review types of spoken language and micro- and macroskills of speaking, then outline numerous tasks for assessing speaking.

BASIC TYPES OF SPEAKING

In Chapter 6, we cited four categories of listening performance assessment tasks. A similar taxonomy emerges for oral production.

1. Imitative. 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. While this is a purely phonetic level of oral production, a number of prosodic, lexical, and grammatical properties of language may be included in the criterion performance. We are interested only in what is traditionally labeled “pronunciation”; no inferences are made about the test-taker’s ability to understand or convey meaning or to participate in an interactive conversation. The only role of listening here is in the short-term storage of a prompt, just long enough to allow the speaker to retain the short stretch of language that must be imitated.

2. Intensive. 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). The speaker must be aware of semantic properties in order to be able to respond, but interaction with an interlocutor or test administrator is minimal at best. 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.

3. Responsive. **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. The stimulus is almost always a spoken prompt (in order to preserve authenticity), with perhaps only one or two follow-up questions or retorts:

A. Mary: Excuse me, do you have the time?
Doug: Yeah. Nine-fifteen.

B. T: What is the most urgent environmental problem today?
S: I would say massive deforestation.

C. Jeff: Hey, Stef, how's it going?
Stef: Not bad, and yourself?
Jeff: I'm good.
Stef: Cool. Okay, gotta go.

4. *Interactive*. The difference between responsive and interactive speaking is in the length and complexity of the interaction, which sometimes includes multiple exchanges and/or multiple participants. Interaction can 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 the three dialogues cited above, A and B were transactional, and C was interpersonal.) 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.

5. *Extensive (monologue)*. 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. Language style is frequently more deliberative (planning is involved) and formal for extensive tasks, but we cannot rule out certain informal monologues such as casually delivered speech (for example, my vacation in the mountains, a recipe for outstanding pasta primavera, recounting the plot of a novel or movie).

MICRO- AND MACROSKILLS OF SPEAKING

In Chapter 6, a list of listening micro- and macroskills enumerated the various components of listening that make up criteria for assessment. A similar list of speaking skills can be drawn up for the same purpose: to serve as a taxonomy of skills from which you will select one or several that will become the objective(s) of an assessment task. The microskills refer to producing the smaller chunks of language such as phonemes, morphemes, words, collocations, and phrasal units. The macroskills imply the speaker's focus on the larger elements: fluency, discourse, function, style, cohesion, nonverbal communication, and strategic options. The micro- and macroskills total roughly 16 different objectives to assess in speaking.

Micro- and macroskills of oral production

Microskills

1. Produce differences among English phonemes and allophonic variants.
2. Produce chunks of language of different lengths.
3. Produce English stress patterns, words in stressed and unstressed positions, rhythmic structure, and intonation contours.
4. Produce reduced forms of words and phrases.
5. Use an adequate number of lexical units (words) to accomplish pragmatic purposes.
6. Produce fluent speech at different rates of delivery.

7. Monitor one's own oral production and use various strategic devices—pauses, fillers, self-corrections, backtracking—to enhance the clarity of the message.
8. Use grammatical word classes (nouns, verbs, etc.), systems (e.g., tense, agreement, pluralization), word order, patterns, rules, and elliptical forms.
9. Produce speech in natural constituents: in appropriate phrases, pause groups, breath groups, and sentence constituents.
10. Express a particular meaning in different grammatical forms.
11. Use cohesive devices in spoken discourse.

Macroskills

12. Appropriately accomplish communicative functions according to situations, participants, and goals.
13. Use appropriate styles, registers, implicature, redundancies, pragmatic conventions, conversation rules, floor-keeping and -yielding, interrupting, and other sociolinguistic features in face-to-face conversations.
14. Convey links and connections between events and communicate such relations as focal and peripheral ideas, events and feelings, new information and given information, generalization and exemplification.
15. Convey facial features, kinesics, body language, and other nonverbal cues along with verbal language.
16. Develop and use a battery of speaking strategies, such as emphasizing key words, rephrasing, providing a context for interpreting the meaning of words, appealing for help, and accurately assessing how well your interlocutor is understanding you.

As you consider designing tasks for assessing spoken language, these skills can act as a checklist of objectives. While the macroskills have the appearance of being more complex than the microskills, both contain ingredients of difficulty, depending on the stage and context of the test-taker.

There is such an array of oral production tasks that a complete treatment is almost impossible within the confines of one chapter in this book. Below is a consideration of the most common techniques with brief allusions to related tasks. As already noted in the introduction to this chapter, consider three important issues as you set out to design tasks:

1. No speaking task is capable of isolating the single skill of oral production. Concurrent involvement of the additional performance of aural comprehension, and possibly reading, is usually necessary.

2. Eliciting the specific criterion you have designated for a task can be tricky because beyond the word level, spoken language offers a number of productive options to test-takers. Make sure your elicitation prompt achieves its aims as closely as possible.

3. Because of the above two characteristics of oral production assessment, it is important to carefully specify scoring procedures for a response so that ultimately you achieve as high a reliability index as possible.

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 Audiolingual 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.

An occasional phonologically focused repetition task is warranted as long as repetition tasks are not allowed to occupy a dominant role in an overall oral production assessment, and as long as you artfully avoid a negative washback effect. Such tasks range from word level to sentence level, usually with each item focusing on a specific phonological criterion. In a simple repetition task, test-takers repeat the stimulus, whether it is a pair of words, a sentence, or perhaps a question (to test for intonation production).

Word repetition task

<i>Test-takers hear:</i>	Repeat after me:	
	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.

Test-takers repeat the stimulus.

A variation on such a task prompts test-takers with a brief written stimulus which they are to read aloud. (In the section below on intensive speaking, some tasks are described in which test-takers read aloud longer texts.) Scoring specifications must be clear in order to avoid reliability breakdowns. A common form of scoring simply indicates a two- or three-point system for each response.

Scoring scale for repetition tasks

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

The longer the stretch of language, the more possibility for error and therefore the more difficult it becomes to assign a point system to the text. In such a case, it may be imperative to score only the criterion of the task. For example, in the sentence "When did they go on vacation?" since the criterion is falling intonation for *wh*-questions, points should be awarded regardless of any mispronunciation.

PHONEPASS® TEST

An example of a popular test that uses imitative (as well as intensive) production tasks is PhonePass, a widely used, commercially available speaking test in many countries. Among a number of speaking tasks on the test, repetition of sentences (of 8 to 12 words) occupies a prominent role. It is remarkable that research on the PhonePass test has supported the construct validity of its repetition tasks not just for a test-taker's phonological ability but also for discourse and overall oral production ability (Townshend et al., 1998; Bernstein et al., 2000; Cascallar & Bernstein, 2000).

The PhonePass test elicits computer-assisted oral production over a telephone. Test-takers read aloud, repeat sentences, say words, and answer questions. With a downloadable test sheet as a reference, test-takers are directed to telephone a designated number and listen for directions. The test has five sections.

*PhonePass® test specifications***Part A:**

Test-takers read aloud selected sentences from among those printed on the test sheet. Examples:

1. Traffic is a huge problem in Southern California.
2. The endless city has no coherent mass transit system.
3. Sharing rides was going to be the solution to rush-hour traffic.
4. Most people still want to drive their own cars, though.

Part B:

Test-takers repeat sentences dictated over the phone. Examples: "Leave town on the next train."

Part C:

Test-takers answer questions with a single word or a short phrase of two or three words. Example: "Would you get water from a bottle or a newspaper?"

Part D:

Test-takers hear three word groups in random order and must link them in a correctly ordered sentence. Example: was reading/my mother/a magazine.

Part E:

Test-takers have 30 seconds to talk about their opinion about some topic that is dictated over the phone. Topics center on family, preferences, and choices.

Scores for the PhonePass test are calculated by a computerized scoring template and reported back to the test-taker within minutes. Six scores are given: an overall score between 20 and 80 and five subscores on the same scale that rate pronunciation, reading fluency, repeat accuracy, repeat fluency, and listening vocabulary.

The tasks on Parts A and B of the PhonePass test do not extend beyond the level of oral reading and imitation. Parts C and D represent intensive speaking (see the next section in this chapter). Section E is used only for experimental data-gathering and does not figure into the scoring. The scoring procedure has been validated against human scoring with extraordinarily high reliabilities and correlation statistics (.94 overall). Further, this ten-minute test correlates with the elaborate Oral Proficiency Interview (OPI, described later in this chapter) at .75, indicating a very high degree of correspondence between the machine-scored PhonePass and the human-scored OPI (Bernstein et al., 2000).

The PhonePass findings could signal an increase in the future use of repetition and read-aloud procedures for the assessment of oral production. Because a test-taker's output is completely controlled, scoring using speech-recognition technology becomes achievable and practical. As researchers uncover the constructs underlying both repetition/read-aloud tasks and oral production in all its complexities, we will have access to more comprehensive explanations of why such simple tasks appear to be reliable and valid indicators of very complex oral production proficiency. Here are some details on the PhonePass test.

PhonePass® test

Producer:	Ordinate Corporation, Menlo Park, CA
Objective:	To test oral production skills of non-native English speakers
Primary market:	Worldwide, primarily in workplace settings where employees require a comprehensible command of spoken English; secondarily in academic settings for placement and evaluation of students

Type:	Computer-assisted telephone operated, with a test sheet
Response modes:	Oral, mostly repetition tasks
Specifications:	(see above)
Time allocations:	Ten minutes
Internet access:	www.ordinate.com

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.

Parts C and D of the PhonePass test fulfill the criteria of intensive tasks as they elicit certain expected forms of language. Antonyms like *high* and *low*, *happy* and *sad* are prompted so that the automated scoring mechanism anticipates only one word. The either/or task of Part D fulfills the same criterion. Intensive tasks may also be described as **limited** response tasks (Madsen, 1983), or **mechanical** tasks (Underhill, 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. Because of the

results of research on the PhonePass test, reading aloud may actually be a surprisingly strong indicator of overall oral production ability.

For many decades, foreign language programs have used reading passages to analyze oral production. Prator's (1972) *Manual of American English Pronunciation* included a "diagnostic passage" of about 150 words that students could read aloud into a tape recorder. Teachers listening to the recording would then rate students on a number of phonological factors (vowels, diphthongs, consonants, consonant clusters, stress, and intonation) by completing a two-page diagnostic checklist on which all errors or questionable items were noted. These checklists ostensibly offered direction to the teacher for emphases in the course to come.

An earlier form of the Test of Spoken English (TSE[®], see below) incorporated one read-aloud passage of about 120 to 130 words with a rating scale for pronunciation and fluency. The following passage is typical:

Read-aloud stimulus, paragraph length

Despite the decrease in size—and, some would say, quality—of our cultural world, there still remain strong differences between the usual British and American writing styles. The question is, how do you get your message across? English prose conveys its most novel ideas as if they were timeless truths, while American writing exaggerates; if you believe 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 and 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.

Such a rating list does not indicate how to gauge *intelligibility*, which is mentioned in both lists. Such slippery terms remind us that oral production scoring, even with the controls that reading aloud offers, is still an inexact science.

Underhill (1987, pp. 77–78) suggested some variations on the task of simply reading a short passage:

- reading a scripted dialogue, with someone else reading the other part
- reading sentences containing minimal pairs, for example:
Try not to heat/hit the pan too much.
The doctor gave me a bill/pill.
- reading information from a table or chart

If reading aloud shows certain practical advantages (predictable output, practicality, reliability in scoring), there are several drawbacks to using this technique for assessing oral production. Reading aloud is somewhat inauthentic in that we seldom read anything aloud to someone else in the real world, with the exception of a parent reading to a child, occasionally sharing a written story with someone, or giving a scripted oral presentation. Also, reading aloud calls on certain specialized oral abilities that may not indicate one's pragmatic ability to communicate orally in face-to-face contexts. You should therefore employ this technique with some caution, and certainly supplement it as an assessment task with other, more communicative procedures.

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.

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.

Such a rating list does not indicate how to gauge *intelligibility*, which is mentioned in both lists. Such slippery terms remind us that oral production scoring, even with the controls that reading aloud offers, is still an inexact science.

Underhill (1987, pp. 77–78) suggested some variations on the task of simply reading a short passage:

- reading a scripted dialogue, with someone else reading the other part
- reading sentences containing minimal pairs, for example:
 - Try not to heat/hit the pan too much.
 - The doctor gave me a bill/pill.
- reading information from a table or chart

If reading aloud shows certain practical advantages (predictable output, practicality, reliability in scoring), there are several drawbacks to using this technique for assessing oral production. Reading aloud is somewhat inauthentic in that we seldom read anything aloud to someone else in the real world, with the exception of a parent reading to a child, occasionally sharing a written story with someone, or giving a scripted oral presentation. Also, reading aloud calls on certain specialized oral abilities that may not indicate one's pragmatic ability to communicate orally in face-to-face contexts. You should therefore employ this technique with some caution, and certainly supplement it as an assessment task with other, more communicative procedures.

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: Hmmm. 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.

An advantage of this technique lies in its moderate control of the output of the test-taker. While individual variations in responses are accepted, the technique taps into a learner's ability to discern expectancies in a conversation and to produce sociolinguistically correct language. One disadvantage of this technique is its reliance on literacy and an ability to transfer easily from written to spoken English. Another disadvantage is the contrived, inauthentic nature of this task: Couldn't the same criterion performance be elicited in a live interview in which an impromptu role-play technique is used?

Perhaps more useful is a whole host of shorter dialogues of two or three lines, each of which aims to elicit a specified target. In the following examples, somewhat unrelated items attempt to elicit the past tense, future tense, *yes/no* question formation, and asking for the time. Again, test-takers see the stimulus in written form.

*Directed response tasks**Test-takers see:*

Interviewer: What did you do last weekend?

Test-taker: _____

Interviewer: What will you do after you graduate from this program?

Test-taker: _____

Test-taker: _____ ?

Interviewer: I was in Japan for two weeks.

Test-taker: _____ ?

Interviewer: It's ten-thirty.

Test-takers respond with appropriate lines.

One could contend that performance on these items is *responsive*, rather than *intensive*. True, the discourse involves responses, but there is a degree of control here that predisposes the test-taker to respond with certain expected forms. Such arguments underscore the fine lines of distinction between and among the selected five categories.

It could also be argued that such techniques are nothing more than a written form of questions that might otherwise (and more appropriately) be part of a standard oral interview. True, but the advantage that the written form offers is to provide a little more time for the test-taker to anticipate an answer, and it begins to remove the potential ambiguity created by aural misunderstanding. It helps to unlock the almost ubiquitous link between listening and speaking performance.

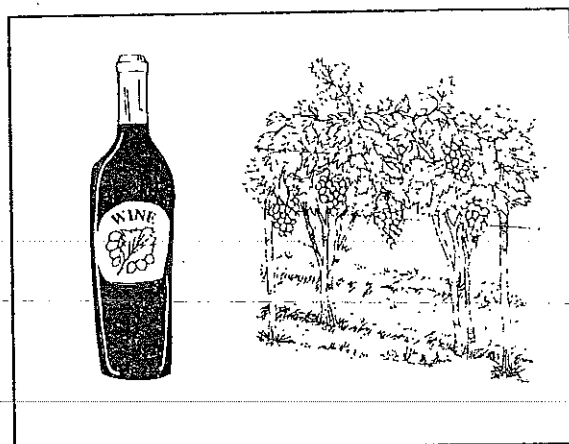
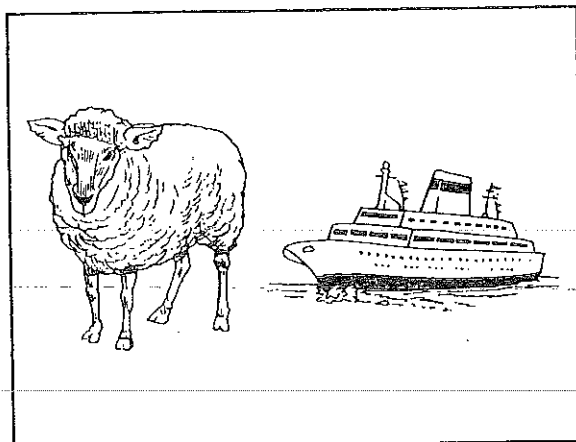
Underhill (1987) describes yet another technique that is useful for controlling the test-taker's output: form-filling, or what I might rename "oral questionnaire." Here the test-taker sees a questionnaire that asks for certain categories of information (personal data, academic information, job experience, etc.) and supplies the information orally.

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:

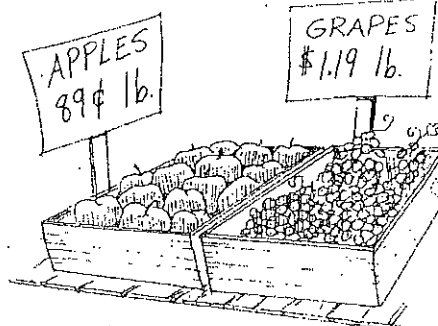
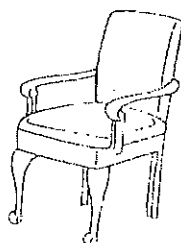
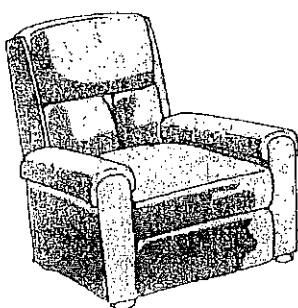


Test-takers hear: [test administrator points to each picture in succession]
What's this?

Grammatical categories may be cued by pictures. In the following sequences, comparatives are elicited:

Picture-cued elicitation of comparatives (Brown & Sahni, 1994, p. 135)

Test-takers see:



Test-takers hear: Use a comparative form to compare these objects.

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?

Notice that a little sense of humor is injected here: the family, bundled up in their winter coats, is looking forward to leaving the wintry scene behind them! A touch of authenticity is added in that almost everyone can identify with looking forward to a vacation on a tropical island.

Assessment of oral production may be stimulated through a more elaborate picture such as the one on the next page, a party scene.

Picture-cued elicitation of nouns, negative responses, numbers, and location
(Brown & Sahni, 1994, p. 116)

Test-takers see:



Test-takers hear:

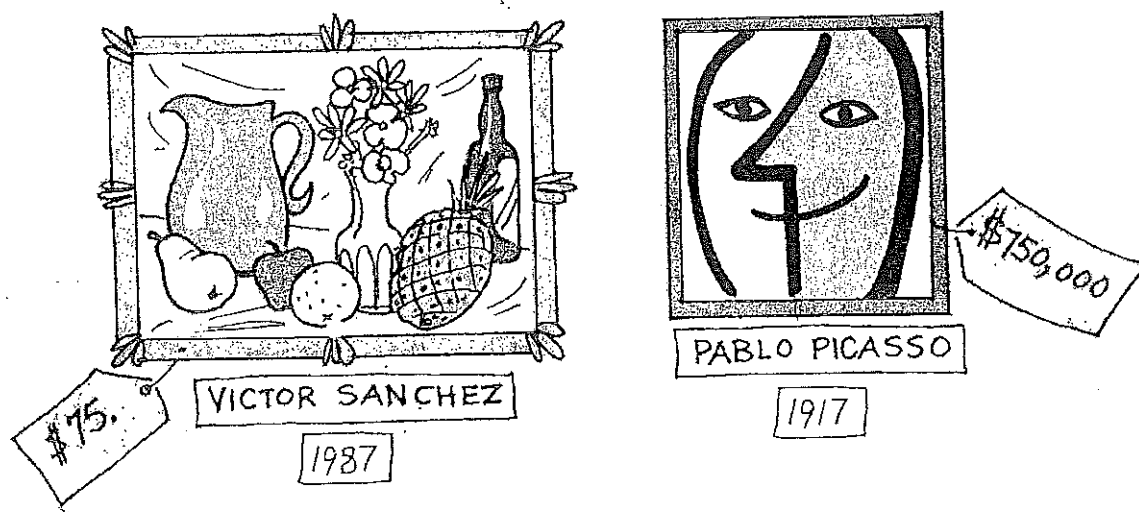
1. [point to the table] What's this?
2. [point to the end table] What's this?
3. [point to several chairs] What are these?
4. [point to the clock] What's that?
5. [point to both lamps] What are those?
6. [point to the table] Is this a chair?
7. [point to the lamps] Are these clocks?
8. [point to the woman standing up] Is she sitting?
9. [point to the whole picture] How many chairs are there?
10. [point to the whole picture] How many women are there?
11. [point to the TV] Where is the TV?
12. [point to the chair beside the lamp] Where is this chair?
13. [point to one person] Describe this person.

In the first five questions, test-takers are asked to orally identify selected vocabulary items. In questions 6-13, assessment of the oral production of negatives, numbers, prepositions, and descriptions of people is elicited.

Moving into more open-ended performance, the following picture asks test-takers not only to identify certain specific information but also to elaborate with their own opinion, to accomplish a persuasive function, and to describe preferences in paintings.

Picture-cued elicitation of responses and description
(Brown & Sahni, 1994, p. 162)

Test-takers see:



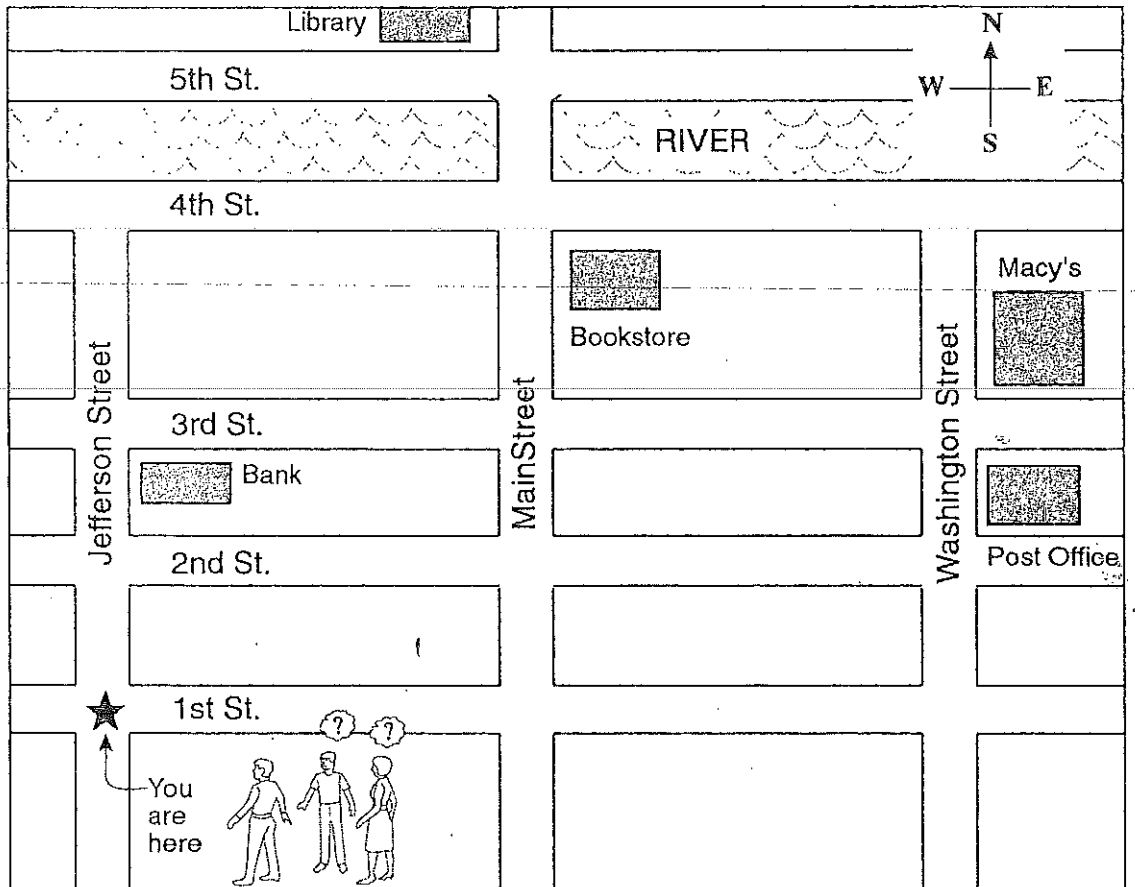
Test-takers hear:

1. [point to the painting on the right] When was this one painted?
[point to both] Which painting is older?
2. [point to the painting on the left] How much does this cost?
Which painting is more expensive?
3. Which painting would you buy? Why?
4. Persuade me to buy it.
5. Describe the kinds of paintings you like [in general].

Maps are another visual stimulus that can be used to assess the language forms needed to give directions and specify locations. In the following example, the test-taker must provide directions to different locations.

Map-cued elicitation of giving directions (Brown & Sahni, 1994, p. 169)

Test-takers see:



Test-takers hear:

You are at First and Jefferson Streets [point to the spot]. People ask you for directions to get to five different places. Listen to their questions, then give directions.

1. Please give me directions to the bank.
2. Please give me directions to Macy's Department Store.
3. How do I get to the post office?
4. Can you tell me where the bookstore is?
5. Please tell me how to get to the library.

Scoring responses on picture-cued intensive speaking tasks varies, depending on the expected performance criteria. The tasks above that asked just for one-word or simple-sentence responses can be evaluated simply as "correct" or "incorrect." The three-point rubric (2, 1, and 0) suggested earlier 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

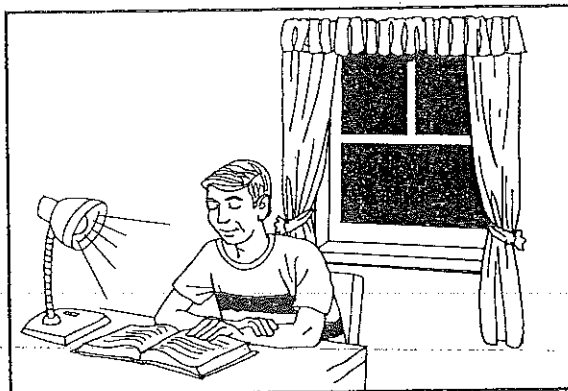
Opinions about paintings, persuasive monologue, and directions on a map create a more complicated problem for scoring. More demand is placed on the test administrator to make calculated judgments, in which case a modified form of a scale such as the one suggested for evaluating interviews (below) could be used:

- grammar
- vocabulary
- comprehension
- fluency
- pronunciation
- task (accomplishing the objective of the elicited task)

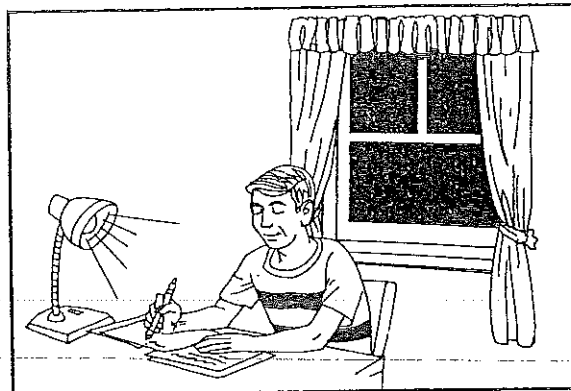
Each category may be scored separately, with an additional composite score that attempts to synthesize overall performance. To attend to so many factors, you will probably need to have an audiotaped recording for multiple listening.

One moderately successful picture-cued technique involves a pairing of two test-takers. They are supplied with a set of four identical sets of numbered pictures, each minimally distinct from the others by one or two factors. One test-taker is directed by a cue card to describe *one* of the four pictures in as few words as possible. The second test-taker must then identify the picture. On the next page is an example of four pictures.

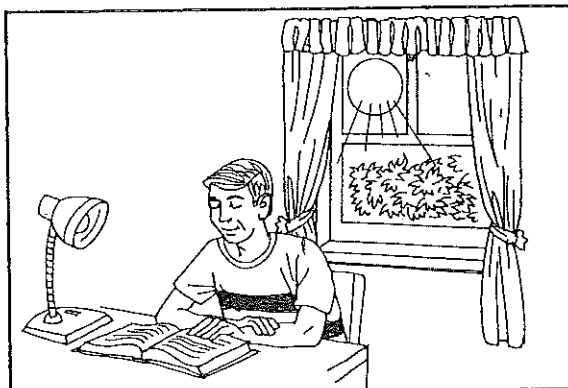
Test-takers see:



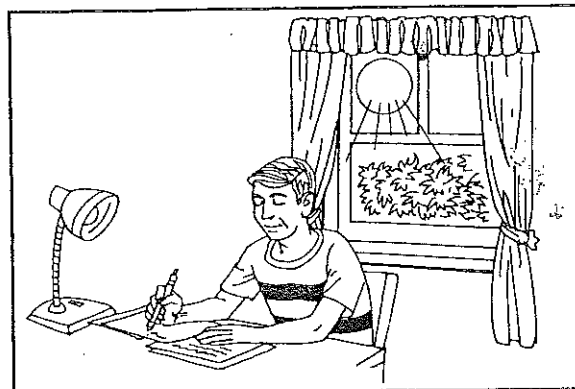
A



B



C



D

Test-taker 1 describes (for example) picture C; test-taker 2 points to the correct picture.

The task here is simple and straightforward and clearly in the intensive category as the test-taker must simply produce the relevant linguistic markers. Yet it is still the task of the test administrator to determine a correctly produced response and a correctly understood response since sources of incorrectness may not be easily pinpointed. If the pictorial stimuli are more complex than the above item, greater burdens are placed on both speaker and listener, with consequently greater difficulty in identifying which committed the error.

Translation (of Limited Stretches of Discourse)

Translation is a part of our tradition in language teaching that we tend to discount or disdain, if only because our current pedagogical stance plays down its importance. Translation methods of teaching are certainly passé in an era of direct approaches to creating communicative classrooms. But we should remember that in countries where English is not the native or prevailing language, translation is a meaningful communicative device in contexts where the English user is called on to be an interpreter. Also, translation is a well-proven communication strategy for learners of a second language.

Under certain constraints, then, it is not far-fetched to suggest translation as a device to check oral production. Instead of offering pictures or written stimuli, the test-taker is given a native language word, phrase, or sentence and is asked to translate it. Conditions may vary from expecting an instant translation of an orally elicited linguistic target to allowing more thinking time before producing a translation of somewhat longer texts, which may optionally be offered to the test-taker in written form. (Translation of extensive texts is discussed at the end of this chapter.) As an assessment procedure, the advantages of translation lie in its control of the output of the test-taker, which of course means that scoring is more easily specified.

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. They can vary from simple questions like "What is this called in English?" to complex questions like "What are the steps governments should take, if any, to stem the rate of deforestation in tropical countries?" The first question is intensive in its purpose; it is a **display question** intended to elicit a predetermined correct response. We have already looked at some of these types of questions in the previous section. Questions at the responsive level tend to be genuine **referential questions** in which the test-taker is given more opportunity to produce meaningful language in response.

ORAL PRODUCTION TESTS

Conversational Exchanges

TYPE 1. The students are given a series of situations and are required to construct sentences on the lines of a certain pattern or group of patterns. It is essential that two or three models be given to the students so that they know exactly what is required.

Examples : Mrs Green lives in a flat. She doesn't like living in a flat and would like to live in a small house with a garden.
(She wishes she lived in a small house with a garden.)

It's raining heavily. Tom and Anna are waiting impatiently at home to set off on their picnic.
(They wish it would stop raining.)

1. Mr Black has a small car but his neighbours all have large cars. He would like a large car, too.
2. Ann hasn't learned how to swim yet, but most of her friends can swim.
3. Tom is waiting for Bill outside the cinema. The show is just about to start, but Bill hasn't arrived yet.

TYPE 2. No model responses.

1. A friend of yours has forgotten where he has put his glasses. He cannot see too well without them. What will you say to him?
2. A waitress has just brought you the bill, but he has totalled it incorrectly. What do you say to her ?

3. You are trying to get to the public library, but you are lost. Ask a police officer the way.

TYPE 3. The students hear a stimulus to which they must respond in any appropriate way.

1. Do you mind if I use your pencil for a moment?
2. What about a game of tennis?
3. Oh, dear, it's raining again. I hope it stops soon.
4. We shan't be late, shall we?

TYPE 4. The stimuli and responses form part of a longer dialogue and the situation is thus developed.

You are on your way to the supermarket. A man comes up and speaks to you.

MAN: Excuse me. I wonder if you can help me at all. I'm looking for a chemist's.

(PAUSE FOR THE STUDENT'S REPLY)

MAN: Thank you. Do you know what time it opens?

(PAUSE FOR THE STUDENT'S REPLY)

MAN: Thanks a lot. Oh, er by the way, is there a phone box near here?

(PAUSE FOR THE STUDENT'S REPLY)

.....

TYPE 5. An incomplete dialogue with prompts which are whispered in the student's ear.

You are at the reception desk of a large hotel. The receptionist turns to address you.

RECEPTIONIST : Can I help you ?

(You want to know if there is a single room available.)

YOU :

RECEPTIONIST : Yes, we have a single room with an attached bathroom.

(Ask the price.)

YOU :

RECEPTIONIST : Thirty-four pounds fifty a night.

(You want to know if this includes breakfast.)

YOU :

RECEPTIONIST : Yes, that's with continental breakfast.

(You have no idea what 'continental breakfast' is.)

YOU :

RECEPTIONIST : It's fruit juice, coffee or tea, and bread rolls.

(The receptionist is speaking too quickly. What do you say?)

YOU :

RECEPTIONIST : Fruit juice, coffee or tea, and bread rolls.

(Book the room for two nights.)

YOU :

RECEPTIONIST : Certainly. Room 216. The porter will take your bag and show you where it is.

(Thank the receptionist.)

YOU :

Using pictures for assessing oral production

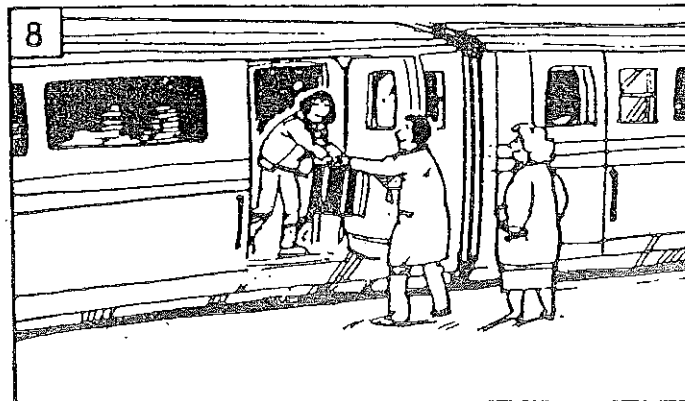
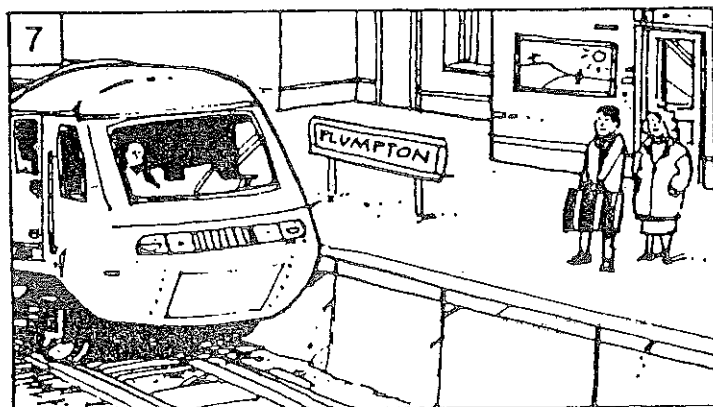
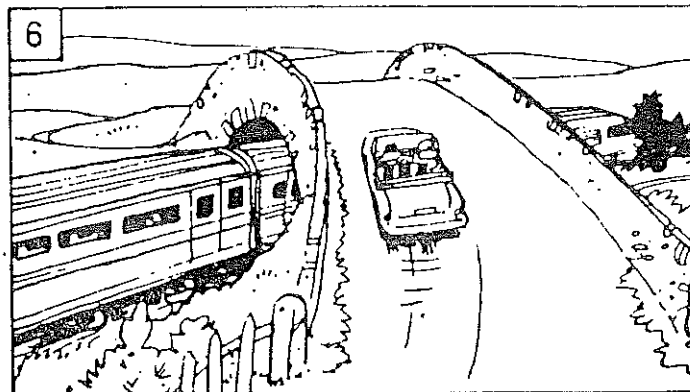
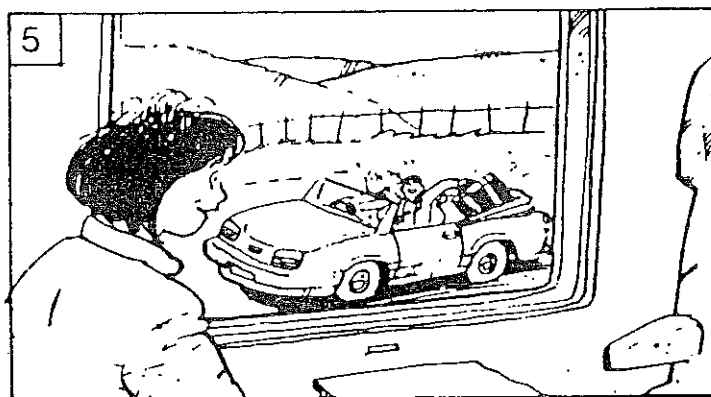
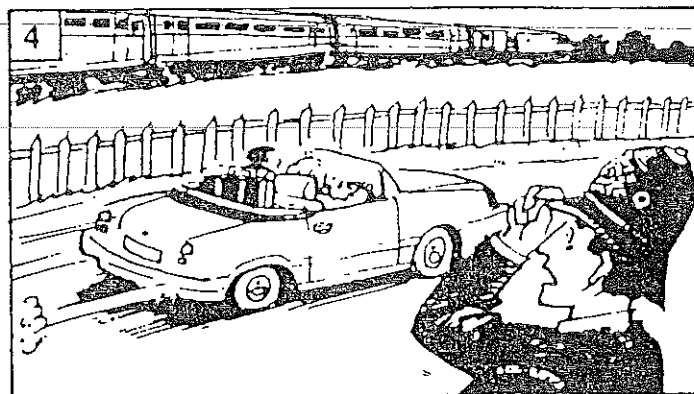
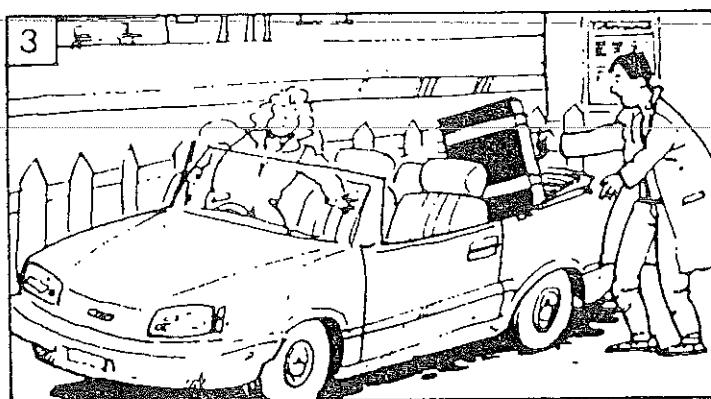
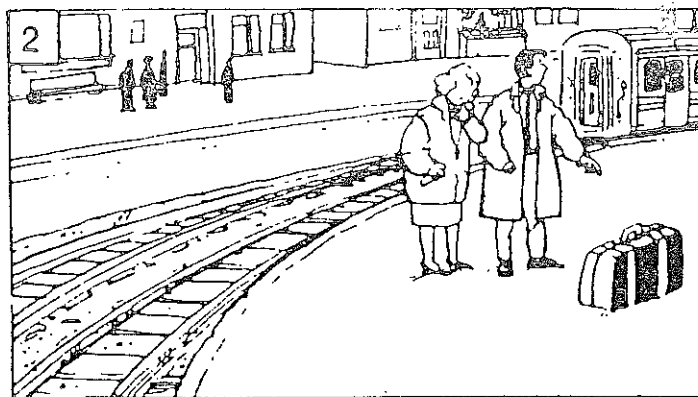
The most traditional way is that the students are given a picture to study for a few minutes; they are then required to describe the picture in a given time (e.g. two or three minutes).

Careful selection of the pictures used for the examination will help in controlling the basic vocabulary required, and may, to some extent, will determine the type of sentence structure that predominates.

If the pictures depict a story or sequence of events, it is useful to give the students one or two sentences as a 'starter', thereby familiarising them with the tense sequencing they should employ.

Example 1 :

Teacher : Last summer Lucy spent a few days with her uncle and aunt in the country. When it was time for her to return home, her uncle and aunt took her to the station. Lucy made a lot of friends and she felt sad on leaving them. She got on the train and waved goodbye to them. Now you continue to tell the story.



Example 2 :

The student and the teacher have pictures in front of them, each picture differing in only one respect from the other four pictures. The student is given a card bearing a letter (A, B, C, D, or D). The teacher cannot see the letter. The student is required to describe the appropriate picture (according to the letter). The teacher then selects a picture according to the description, and checks the card.



SCORING

Consider the following.

The teacher : What are you going to do this weekend ?

Student A : I'm quite well, thank you.

Student B : I go to fish. I fish in river near the big wood.

A' reply is perfectly **correct**, but it is nevertheless quite **inappropriate**.

The student simply hasn't heard or understood the question correctly.

On the other hand, B' reply shows a real attempt the question but unfortunately contain several errors.

The following marking scheme (using a 6-point scale) is given as just one example of a number of such schemes in present-day use.

Rating	Ability to communicate orally
6	Excellent: on a par with an educated native speaker. Completely at ease in his use of English on all topics discussed.
5	Very good: although he cannot be mistaken for a native speaker, he expresses himself quite clearly. He experiences little difficulty in understanding English, and there is no strain at all in communicating with him.
4	Satisfactory verbal communication causing little difficulty for native speakers. He makes a limited number of errors of grammar, lexis and pronunciation but he is still at ease in communicating on everyday subjects. He may have to correct himself and repattern his utterance on occasions, but there is little difficulty in understanding him.
3	Although verbal communication is usually fairly satisfactory, the native speaker may occasionally experience some difficulty in communicating with him. Repetition, re-phrasing and re-patterning are sometimes necessary; ordinary native speakers might find it difficult to communicate.
2	Much difficulty experienced by native speakers unaccustomed to 'foreign' English. His own understanding is severely limited, but communication on everyday topics is possible. Large number of errors of phonology, grammar and lexis.
1	Extreme difficulty in communication on any subject. Failure to understand adequately and to make himself understood.

The following is an example of a teacher's rating scale for the lower intermediate level.

Accuracy	Fluency	Comprehensibility
6 Pronunciation is only very slightly influenced by the mother-tongue. Two or three minor grammatical and lexical errors.	Speaks without too great an effort with a fairly wide range of expression. Searches for words occasionally but only one or two unnatural pauses.	Easy for the listener to understand the speaker's intention and general meaning. Very few interruptions or clarifications required.
5 Pronunciation is slightly influenced by the mother-tongue. A few minor grammatical and lexical errors but most utterances are correct.	Has to make an effort at times to search for words. Nevertheless, smooth delivery on the whole and only a few unnatural pauses.	The speaker's intention and general meaning are fairly clear. A few interruptions by the listener for the sake of clarification are necessary.
4 Pronunciation is still moderately influenced by the mother-tongue but no serious phonological errors. A few grammatical and lexical errors but only one or two major errors causing confusion.	Although he has to make an effort and search for words, there are not too many unnatural pauses. Fairly smooth delivery mostly. Occasionally fragmentary but succeeds in conveying the general meaning. Fair range of expression.	Most of what the speaker says is easy to follow. His intention is always clear but several interruptions are necessary to help him to convey the message or to seek clarification.
3 Pronunciation is influenced by the mother-tongue but only a few serious phonological errors. Several grammatical and lexical errors, some of which cause confusion.	Has to make an effort for much of the time. Often has to search for the desired meaning. Rather halting delivery and fragmentary. Range of expression often limited.	The listener can understand a lot of what is said, but he must constantly seek clarification. Cannot understand many of the speaker's more complex or longer sentences.
2 Pronunciation seriously influenced by the mother-tongue with errors causing a breakdown in communication. Many 'basic' grammatical and lexical errors.	Long pauses while he searches for the desired meaning. Frequently fragmentary and halting delivery. Almost gives up making the effort at times. Limited range of expression.	Only small bits (usually short sentences and phrases) can be understood – and then with considerable effort by someone who is used to listening to the speaker.
1 Serious pronunciation errors as well as many 'basic' grammatical and lexical errors. No evidence of having mastered any of the language skills and areas practised in the course.	Full of long and unnatural pauses. Very halting and fragmentary delivery. At times gives up making the effort. Very limited range of expression.	Hardly anything of what is said can be understood. Even when the listener makes a great effort or interrupts, the speaker is unable to clarify anything he seems to have said.

