# 10 Testing oral ability

The assumption is made in this chapter that the objective of teaching spoken language is the development of the ability to interact successfully in that language, and that this involves comprehension as well as production. It is also assumed that at the earliest stages of learning formal testing of this ability will not be called for, informal observation providing any diagnostic information that is needed.

The basic problem in testing oral ability is essentially the same as for testing writing.

1. We want to set tasks that form a representative sample of the population of oral tasks that we expect candidates to be able to perform.
2. The tasks should elicit behaviour which truly represents the candidates' ability.
3. The samples of behaviour can and will be scored validly and reliably.

Following the pattern of the previous chapter, we shall deal with each of these in turn.

## Representative tasks

### *Specify all possible content*

We will begin by looking at the specified content of the Cambridge CCSE Test of Oral Interaction, covering all four levels at which a certificate is awarded.

**Operations**[1]

*Expressing:* likes, dislikes, preferences, agreement/disagreement, requirements, opinions, comment, attitude, confirmation, complaints, reasons, justifications, comparisons
*Directing:* instructing, persuading, advising, prioritising
*Describing:* actions, events, objects, people, processes

*Eliciting:* information, directions, clarification, help
*Narration:* sequence of events
*Reporting:* description, comment, decisions and choices

**Types of text** Discussion

**Addressees** 'Interlocutor' (teacher from candidate's school) and one fellow candidate

**Topics** Unspecified

**Dialect, Accent and Style** also unspecified

It can be seen that the content specifications are similar to those for the Test of Writing. They may be compared with those for a test with which I have been concerned. The categorisation of the operations (here referred to as skills) is based on Bygate (1987).

**Skills**

***Informational skills***
Candidates should be able to:

- provide personal information
- provide non-personal information
- describe sequence of events (narrate)
- give instructions
- make comparisons
- give explanations
- present an argument
- provide required information
- express need
- express requirements
- elicit help
- seek permission
- apologise
- elaborate an idea
- express opinions
- justify opinions
- complain
- speculate
- analyse
- make excuses
- paraphrase
- summarise (what they have said)
- make suggestions

- express preferences
- draw conclusions
- make comments
- indicate attitude

***Interactional skills***
Candidates should be able to:
- express purpose
- recognise other speakers' purpose
- express agreement
- express disagreement
- elicit opinions
- elicit information
- question assertions made by other speakers
- modify statements or comments
- justify or support statements or opinions of other speakers
- attempt to persuade others
- repair breakdowns in interaction
- check that they understand or have been understood correctly
- establish common ground
- elicit clarification
- respond to requests for clarification
- correct themselves or others
- indicate understanding (or failure to understand)
- indicate uncertainty

***Skills in managing interactions***
Candidates should be able to:
- initiate interactions
- change the topic of an interaction
- share the responsibility for the development of an interaction
- take their turn in an interaction
- give turns to other speakers
- come to a decision
- end an interaction

**Types of text**
- Presentation (monologue)
- Discussion
- Conversation
- Service encounter
- Interview

**Other speakers (addressees)**
- may be of equal or higher status
- may be known or unknown

**Topics** Topics which are familiar and interesting to the candidates

**Dialect** Standard British English or Standard American English

**Accent** RP, Standard American

**Style** Formal and informal

**Vocabulary range** Non-technical except as the result of preparation for a presentation

**Rate of speech** Will vary according to task

It can be seen that this second set of content specifications is rather fuller than the first. What is more, splitting the skills into three categories (informational, interactional, and management), as it does, should help in creating tasks which will elicit a representative sample of each. In my view, the greater the detail in the specification of content, the more valid the test is likely to be. Readers may wish to select elements from the two sets of specifications for their own purposes.

### *Include a representative sample of the specified content when setting tasks*

Any one oral test should sample from the full specified range. The reasons for doing this are the same as those given in the previous chapter. Let us look at the materials for a recent Level 4 CCSE test. The test has two sections. In the first section a candidate talks with a teacher from their own institution. In the second they talk with a fellow student, and after some time the teacher joins in their discussion[2].

It is interesting to try to predict which of the functions listed in the specifications would be elicited by these tasks. You might want to attempt to do this before reading any further. Looking at them myself, I thought that in performing the tasks the speakers were quite likely to express opinions, likes and dislikes, preferences, reasons, justifications. They might also describe, narrate or report, depending perhaps on the nature of the justification they provide for their opinions and preferences. It came as a surprise to me therefore to read in the Examiners' Report for this test that the aim of the first task was to elicit 'describing, explaining and justifying', and that of the second was to elicit 'exchanging opinions and justifying'. But it does allow me to make two related

**Section I**

1 You have 5 minutes to read the task and **think** about what you want to say.

2 If there is anything which you don't understand, please ask the teacher who is with you.

3 You can make a few notes if you want to. The examiner will not look at them.

4 After this 5 minute preparation time, you will go into the exam room and talk about the subject with a teacher. The examiner will listen.

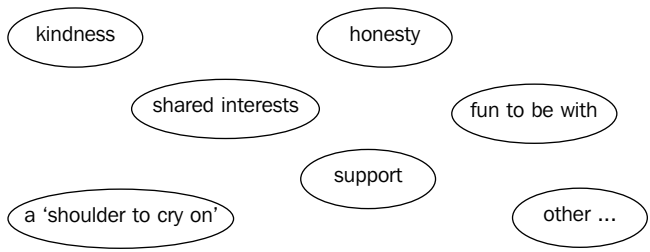
---

**TASK 3**

What makes a good friend?

**You are going to talk to the teacher about what you value in your friends.**

**Look at the suggestions below:**




**Do you think it's better to have one or two really close friends, or a wider circle of less close friends?**

**What are the qualities in yourself that you think your friends value?**

**There is an English saying, "Blood is thicker than water", meaning that family relationships are more important/reliable than relationships with friends. Do you agree with this?**

points. The first is that, unless the tasks are extremely restrictive (which they are not in the CCSE test), it is not possible to predict all the operations which will be performed in an interactive oral test. The second point is that, even where quite specific tasks are set, as in the present case, the 'interlocutor' can have a considerable influence on the content of an oral test. Interviewers therefore need to be well trained and always aware of the need to elicit a representative sample of the operations listed in the specifications.

**Section II**

1 You have 5 minutes to read the task and **think** about what you want to say.

2 If there is anything which you don't understand, please ask the teacher who is with you. **DON'T start talking with your partner yet.**

3 You can make a few notes if you want to. The examiner will not look at them.

4 After this 5 minute preparation time, you will go into the exam room with your partner.

5 The teacher will start your discussion with you and will then leave the room. He or she will join your conversation later for a further 5 minutes. The examiner will listen.

---

**TASK 1**

Whether you have a mobile phone or not, many people have opinions about them.

**Look at the statements below. Tick (✓) the ones <u>you</u> agree with.**

☐ "I hate it when phones ring at the theatre or cinema."

☐ "If you have a mobile phone you never feel alone."

☐ "It's really dangerous to drive and phone at the same time."

☐ "I feel safer with a mobile phone."

☐ "I hate them – people look stupid walking around talking on the phone!"

**Exchange your ideas about mobile phones with your partner. Talk about the reasons why people have them. What advantages do they have over conventional phones? Are there any disadvantages?**

**When the teacher returns, tell him/her about your discussion. S/he will then ask you what limits (if any) should be put on when and where mobile phones can be used.**

**In what ways, for better or worse, is technology changing how we communicate with each other? What about future developments?**

## Elicit a valid sample of oral ability

### *Choose appropriate techniques*

Three general formats are presented here: interview; interaction with fellow candidates; responses to audio- or video-recorded stimuli.

#### Format 1 Interview

Perhaps the most common format for the testing of oral interaction is the interview. In its traditional form, however, it has at least one potentially serious drawback. The relationship between the tester and the candidate is usually such that the candidate speaks as to a superior and is unwilling to take the initiative. As a result, only one style of speech is elicited, and many functions (such as asking for information) are not represented in the candidate's performance. It is possible, however, to get round this problem by introducing a variety of elicitation techniques into the interview situation.

Useful techniques are:

#### Questions and requests for information

Yes/No questions should generally be avoided, except perhaps at the very beginning of the interview, while the candidate is still warming up. Performance of various operations (of the kind listed in the two sets of specifications above) can be elicited through requests of the kind:

> 'Can you explain to me how/why ...?' and
>
> 'Can you tell me what you think of …?'

**Requests for elaboration:** such as *What exactly do you mean?*, *Can you explain that in a little more detail?*, *What would be a good example of that? Tell me more.*

**Appearing not to understand:** This is most appropriate where the interviewer really isn't sure of what the candidate means but can also be used simply in order to see if the candidate can cope with being misunderstood. The interviewer may say, for example, *I'm sorry, but I don't quite follow you.*

**Invitation to ask questions:** *Is there anything you'd like to ask me?*

**Interruption:** To see how the candidate deals with this.

**Abrupt change of topic:** To see how the candidate deals with this.

### Pictures

Single pictures are particularly useful for eliciting descriptions. Series of pictures (or video sequences) form a natural basis for narration (the series of pictures on page 92 for example).

### Role play

Candidates can be asked to assume a role in a particular situation. This allows the ready elicitation of other language functions. There can be a series of brief items, such as:

> A friend invites you to a party on an evening when you want to stay at home and watch the last episode of a television serial. Thank the friend (played by the tester) and refuse politely.

Or there can be a more protracted exchange:

> You want your mother (played by the tester) to increase your pocket money. She is resistant to the idea. Try to make her change her mind.
>
> You want to fly from London to Paris on 13 March, returning a week later. Get all the information that you need in order to choose your flights from the travel agent (played by the tester).

In my experience, however, where the aim is to elicit 'natural' language and an attempt has been made to get the candidates to forget, to some extent at least, that they are being tested, role play can destroy this illusion. I have found that some candidates, rather than responding to the situation as if it were one they were actually facing, will resort to uttering half remembered snatches of exchanges once learned by rote.

### Interpreting

It is not intended that candidates should be able to act as interpreters (unless that is specified). However, simple interpreting tasks can test both production and comprehension in a controlled way. If there are two testers, one of the testers acts as a monolingual speaker of the candidate's native language, the other as a monolingual speaker of the language being tested. Situations of the following kind can be set up:

> The native language speaker wants to invite a foreign visitor to his or her home for a meal. The candidate has to convey the invitation and act as an interpreter for the subsequent exchange.

Comprehension can be assessed when the candidate attempts to convey what the visitor is saying, and indeed unless some such device is used, it is difficult to obtain sufficient information on candidates' powers of

comprehension. Production is tested when the candidate tries to convey the meaning of what the native speaker says.

### Prepared monologue

In the first edition of this book I said that I did not recommend prepared monologues as a means of assessing candidates' oral ability. This was because I knew that the technique was frequently misused. What I should have said is that it should only be used where the ability to make prepared presentations is something that the candidates will need. Thus it could be appropriate in a proficiency test for teaching assistants, or in an achievement test where the ability to make presentations is an objective of the course.

### Reading aloud

This is another technique the use of which I discouraged in the first edition, pointing out that there are significant differences amongst native speakers in the ability to read aloud, and that interference between the reading and the speaking skills was inevitable. But, if that ability is needed or its development has been a course objective, use of the technique may be justified.

## Format 2 Interaction with fellow candidates

An advantage of having candidates interacting with each other is that it should elicit language that is appropriate to exchanges between equals, which may well be called for in the test specifications. It may also elicit better performance, inasmuch as the candidates may feel more confident than when dealing with a dominant, seemingly omniscient interviewer.

There is a problem, however. The performance of one candidate is likely to be affected by that of the others. For example, an assertive and insensitive candidate may dominate and not allow another candidate to show what he or she can do. If interaction with fellow candidates is to take place, the pairs should be carefully matched whenever possible. In general, I would advise against having more than two candidates interacting, as with larger numbers the chance of a diffident candidate failing to show their ability increases.

Possible techniques are:

### Discussion

An obvious technique is to set a task which demands discussion between the two candidates, as in the Test of Oral Interaction above. Tasks may require the candidates to go beyond discussion and, for example, take a decision.

### Role play

Role play can be carried out by two candidates with the tester as an observer. For some roles this may be more natural than if the tester were involved. It may, for example, be difficult to imagine the tester as 'a friend'. However, I believe that the doubts about role play expressed above still apply.

## Format 3 Responses to audio- or video-recordings

Uniformity of elicitation procedures can be achieved through presenting all candidates with the same computer generated or audio-/video-recorded stimuli (to which the candidates themselves respond into a microphone). This format, often described as 'semi-direct', ought to promote reliability. It can also be economical where a language laboratory is available, since large numbers of candidates can be tested at the same time. The obvious disadvantage of this format is its inflexibility: there is no way of following up candidates' responses.

A good source of techniques is the ARELS (Association of Recognised English Language Schools) Examination in Spoken English and Comprehension. These include:

### Described situations

For example:

> You are walking through town one day and you meet two friends who you were sure had gone to live in the USA. What do you say?

### Remarks in isolation to respond to

For example:

> The candidate hears, 'I'm afraid I haven't managed to fix that cassette player of yours yet. Sorry.'
>
> or 'There's a good film on TV tonight.'

### Simulated conversation

For example:

> The candidate is given information about a play which they are supposed to want to see, but not by themselves. The candidate is told to talk to a friend, Ann, on the telephone, and ask her to go to the theatre and answer her questions. The candidate hears:
>
> Ann: Hello. What can I do for you?

Ann: Hold on a moment. What's the name of the play, and who's it by?

Ann: Never heard of it. When's it on exactly?

Ann: Sorry to mention it, but I hope it isn't too expensive.

Ann: Well which night do you want to go, and how much would you like to pay?

Ann: OK. That's all right. It'll make a nice evening out. 'Bye.

Note that although what Ann says is scripted, the style of speech is appropriately informal. For all of the above, an indication is given to candidates of the time available (for example ten seconds) in which to respond. Note, too, that there is room for confusion towards the end of the exchange if the candidate does not say that there are different priced tickets. This is something to be avoided.

The Test of Spoken English (TSE), developed by Educational Testing Services, uses the same elicitation techniques that are found in interviews. In the sample test found in the Standard-setting Kit:

> Candidates see a simple town plan and are asked for (a) recommendation for a visit to one of the buildings, with reasons; (b) directions to the movie theatre; (c) a summary of a favourite movie and their reasons for liking it.
>
> Candidates are given a series of pictures in which a man sits on a recently painted park bench and asked to (a) narrate the story (b) say how the accident could have been avoided (c) imagine that the accident has happened to them and they must persuade the dry cleaners to clean their suit the same day (d) state the advantages and disadvantages of newspapers and television as sources of news (the man in the pictures reads a newspaper on the park bench!).
>
> Candidates are asked to talk about the desirability or otherwise of keeping animals in zoos, define a key term in their field of study, describe the information given in a graph and discuss its implications.
>
> Candidates are given printed information about a trip which has had some handwritten amendments made to it. They must make a presentation to the group of people who are going on the trip, explaining the changes.
>
> Candidates are told how long they have to study the information they are given and how long they are expected to speak for.

Both the ARELS test and the TSE provide useful models for anyone interested in developing tape mediated speaking tests. Notice, however, that the TWE does not make any real attempt to assess interactive skills.

### *Plan and structure the testing carefully*

1. Make the oral test as long as is feasible. It is unlikely that much reliable information can be obtained in less than about 15 minutes, while 30 minutes can probably provide all the information necessary for most purposes. As part of a placement test, however, a five- or ten-minute interview should be sufficient to prevent gross errors in assigning students to classes.
2. Plan the test carefully. While one of the advantages of individual oral testing is the way in which procedures can be adapted in response to a candidate's performance, the tester should nevertheless have some pattern to follow. It is a mistake to begin, for example, an interview with no more than a general idea of the course that it might take. Simple plans of the kind illustrated below can be made and consulted unobtrusively during the interview

**INTRO:** Name, etc.
How did you get here today? traffic problems?

**School:** position, class sizes, children
Typical school day; school holidays
3 pieces of advice to new teachers
Examinations and tests
Tell me about typical errors in English
How do you teach ... present perfect v. past tense
future time reference
conditionals

**What if...** you hadn't become a teacher
**...** you were offered promotion

**INTERPRETING:** How do I get onto the Internet?
How do I find out about the cheapest flights to Europe?

**NEWSPAPER:** (look at the headlines)

**EXPLAIN IDIOMS:** For example, 'Once in a blue moon' or 'See the light'

3. Give the candidate as many 'fresh starts' as possible. This means a number of things. First, if possible and if appropriate, more than one format should be used. Secondly, again if possible, it is desirable for

candidates to interact with more than one tester. Thirdly, within a format there should be as many separate 'items' as possible. Particularly if a candidate gets into difficulty, not too much time should be spent on one particular function or topic. At the same time, candidates should not be discouraged from making a second attempt to express what they want to say, possibly in different words.

4. Use a second tester for interviews. Because of the difficulty of conducting an interview and of keeping track of the candidate's performance, it is very helpful to have a second tester present. This person can not only give more attention to how the candidate is performing but can also elicit performance which they think is necessary in order to come to a reliable judgement. The interpretation task suggested earlier needs the co-operation of a second tester.
5. Set only tasks and topics that would be expected to cause candidates no difficulty in their own language.
6. Carry out the interview in a quiet room with good acoustics.
7. Put candidates at their ease so that they can show what they are capable of. Individual oral tests will always be particularly stressful for candidates. It is important to be pleasant and reassuring throughout, showing interest in what the candidate says through both verbal and non-verbal signals. It is especially important to make the initial stages of the test well within the capacities of all reasonable candidates. Interviews, for example, can begin with straightforward requests for personal (but not too personal) details, remarks about the weather, and so on.

Testers should avoid constantly reminding candidates that they are being assessed. In particular they should not be seen to make notes on the candidates' performance during the interview or other activity. For the same reason, transitions between topics and between techniques should be made as natural as possible. The interview should be ended at a level at which the candidate clearly feels comfortable, thus leaving him or her with a sense of accomplishment.

8. Collect enough relevant information. If the purpose of the test is to determine whether a candidate can perform at a certain predetermined level, then, after an initial easy introduction, the test should be carried out at that level. If it becomes apparent that a candidate is clearly very weak and has no chance of reaching the criterion level, then an interview should be brought gently to a close, since nothing will be learned from subjecting her or him to a longer ordeal. Where, on the other hand, the purpose of the test is to see what level the candidate is at, in an interview the tester has to begin by guessing what this level is on the basis of early responses. The

interview is then conducted at that level, either providing confirmatory evidence or revealing that the initial guess is inaccurate. In the latter case the level is shifted up or down until it becomes clear what the candidate's level is. A second tester, whose main role is to assess the candidate's performance, can elicit responses at a different level if it is suspected that the principal interviewer may be mistaken.

9. Do not talk too much. There is an unfortunate tendency for interviewers to talk too much, not giving enough talking time to candidates. Avoid the temptation to make lengthy or repeated explanations of something that the candidate has misunderstood.
10. Select interviewers carefully and train them. Successful interviewing is by no means easy and not everyone has great aptitude for it. Interviewers need to be sympathetic and flexible characters, with a good command of the language themselves. But even the most apt need training. What follows is the outline of a possible four-stage training programme for interviewers, where interviewing is carried out as recommended above, with two interviewers.

### Stage 1 Background and overview

- Trainees are given background on the interview.
- Trainees are given a copy of the handbook and taken through its contents.
- The structure of the interview is described.
- A video of a typical interview is shown.
- Trainees are asked to study the handbook before the second stage of the training.

### Stage 2 Assigning candidates to levels

- Queries arising from reading the handbook are answered.
- A set of calibrated videos is shown.
- After each video, trainees are asked to write down the levels to which they assign the candidate according to the level descriptions and the analytic scale, and to complete a questionnaire on the task. A discussion follows.
- All papers completed by trainees during this stage are kept as a record of their performance.

### Stage 3 Conducting interviews

- Pairs of trainees conduct interviews, which are videoed.
- The other trainees watch the interview on a monitor in another room.
- After each interview, all trainees assign the candidate to a level and complete a questionnaire. These are then discussed.
- Each trainee will complete 6 interviews.

Stage 4 Assessment

- Procedures will be as in Stage 3, except that the performance of trainees will not be watched by other trainees. Nor will there be any discussion after each interview.

## Ensure valid and reliable scoring

### *Create appropriate scales for scoring*

As was said for tests of writing in the previous chapter, rating scales may be holistic or analytic. The advantages and disadvantages of the two approaches have already been discussed in the previous chapter. We begin by looking at the degree of skill that Level 3 candidates for the CCSE Test of Oral Interaction are required to show. These will have been applied to candidates performing the tasks presented above.

| | |
|---|---|
| ACCURACY | Pronunciation must be clearly intelligible even if some influences from L1 remain. Grammatical/lexical accuracy is high though grammatical errors which do not impede communication are acceptable. |
| APPROPRIACY | The use of language must be generally appropriate to function and to context. The intention of the speaker must be clear and unambiguous. |
| RANGE | A wide range of language must be available to the candidate. Any specific items which cause difficulties can be smoothly substituted or avoided. |
| FLEXIBILITY | There must be consistent evidence of the ability to 'turn-take' in a conversation and to adapt to new topics or changes of direction. |
| SIZE | Must be capable of making lengthy and complex contributions where appropriate. Should be able to expand and develop ideas with minimal help from the Interlocutor. |

Notice that certain elements in these descriptions of degree of skill (such as 'ability to turn-take') could be placed in the content section of the specifications. As long as such elements are taken into account in constructing the tasks (and they are in the CCSE test) this would not seem to be a problem. The CCSE differs from the ILR descriptors below in that the CCSE does specify functions separately.

The ILR speaking levels go from 0 (zero) to 5 (native speaker like), with a plus indicating a level intermediate between two 'whole number' levels. Levels 2, 2+ and 3 follow.

**Speaking 2 (Limited Working Proficiency)**

Able to satisfy routine social demands and limited work requirements. Can handle routine work-related interactions that are limited in scope. In more complex and sophisticated work-related tasks, language usage generally disturbs the native speaker. Can handle with confidence, but not with facility, most normal, high-frequency social conversational situations including extensive, but casual conversations about current events, as well as work, family, and autobiographical information. The individual can get the gist of most everyday conversations but has some difficulty understanding native speakers in situations that require specialized or sophisticated knowledge. The individual's utterances are minimally cohesive. Linguistic structure is usually not very elaborate and not thoroughly controlled; errors are frequent. Vocabulary use is appropriate for high-frequency utterances, but unusual or imprecise elsewhere.

***Examples:*** While these interactions will vary widely from individual to individual, the individual can typically ask and answer predictable questions in the workplace and give straightforward instructions to subordinates. Additionally, the individual can participate in personal and accommodation-type interactions with elaboration and facility; that is, can give and understand complicated, detailed, and extensive directions and make non-routine changes in travel and accommodation arrangements. Simple structures and basic grammatical relations are typically controlled; however, there are areas of weakness. In the commonly taught languages, these may be simple markings such as plurals, articles, linking words, and negatives or more complex structures such as tense/aspect usage, case morphology, passive constructions, word order, and embedding.

**Speaking 2+ (Limited Working Proficiency, Plus)**

Able to satisfy most work requirements with language usage that is often, but not always, acceptable and effective. The individual shows considerable ability to communicate effectively on topics relating to particular interests and special fields of competence.

Often shows a high degree of fluency and ease of speech, yet when under tension or pressure, the ability to use the language effectively may deteriorate. Comprehension of normal native speech is typically nearly complete. The individual may miss cultural and local references and may require a native speaker to adjust to his/her limitations in some ways. Native speakers often perceive the individual's speech to contain awkward or inaccurate phrasing of ideas, mistaken time, space, and person references, or to be in some way inappropriate, if not strictly incorrect.

***Examples:*** Typically the individual can participate in most social, formal, and informal interactions; but limitations either in range of contexts, types of tasks, or level of accuracy hinder effectiveness. The individual may be ill at ease with the use of the language either in social interaction or in speaking at length in professional contexts. He/she is generally strong in either structural precision or vocabulary, but not in both. Weakness or unevenness in one of the foregoing, or in pronunciation, occasionally results in miscommunication. Normally controls, but cannot always easily produce, general vocabulary. Discourse is often incohesive.

**Speaking 3 (General Professional Proficiency)**

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. Nevertheless, the individual's limitations generally restrict the professional contexts of language use to matters of shared knowledge and/or international convention. Discourse is cohesive. The individual uses the language acceptably, but with some noticeable imperfections; yet, errors virtually never interfere with understanding and rarely disturb the native speaker. The individual can effectively combine structure and vocabulary to convey his/her meaning accurately. The individual speaks readily and fills pauses suitably. In face-to-face conversation with natives speaking the standard dialect at a normal rate of speech, comprehension is quite complete. Although cultural references, proverbs, and the implications of nuances and idiom may not be fully understood, the individual can easily repair the conversation. Pronunciation may be obviously foreign. Individual sounds are accurate; but stress, intonation, and pitch control may be faulty.

***Examples:*** Can typically discuss particular interests and special fields of competence with reasonable ease. Can use the language as part of normal professional duties such as answering objections, clarifying points, justifying decisions, understanding the essence of challenges, stating and defending policy, conducting meetings, delivering briefings, or other extended and elaborate informative monologues. Can reliably elicit information and informed opinion from native speakers. Structural inaccuracy is rarely the major cause of misunderstanding. Use of structural devices is flexible and elaborate. Without searching for words or phrases, individual uses the language clearly and relatively naturally to elaborate concepts freely and make ideas easily understandable to native speakers. Errors occur in low-frequency and highly complex structures.

It was said that holistic and analytic scales can be used as a check on each other. An example of this in oral testing is the American FSI (Foreign Service Institute) interview procedure[3], which requires the two testers concerned in each interview both to assign candidates to a level holistically and to rate them on a six-point scale for each of the following: accent, grammar, vocabulary, fluency, comprehension. These ratings are then weighted and totalled. The resultant score is then looked up in a table which converts scores into the holistically described levels. The converted score should give the same level as the one to which the candidate was first assigned. If not, the testers will have to reconsider whether their first assignments were correct. The weightings and the conversion tables are based on research which revealed a very high level of agreement between holistic and analytic scoring. Having used this system myself when testing bank staff, I can attest to its efficacy. For the reader's interest I reproduce the rating scales and the weighting table. It must be remembered, however, that these were developed for a particular purpose and should not be expected to work well in a significantly different situation without modification. It is perhaps also worth mentioning that the use of a native-speaker standard against which to judge performance has recently come in for criticism in some language testing circles.

**Proficiency Descriptions**

**Accent**

1. Pronunciation frequently unintelligible.
2. Frequent gross errors and a very heavy accent make understanding difficult, require frequent repetition.
3. "Foreign accent" requires concentrated listening, and mispronunciations lead to occasional misunderstanding and apparent errors in grammar or vocabulary.
4. Marked "foreign accent" and occasional mispronunciations which do not interfere with understanding.
5. No conspicuous mispronunciations, but would not be taken for a native speaker.
6. Native pronunciation, with no trace of "foreign accent."

**Grammar**

1. Grammar almost entirely inaccurate except in stock phrases.
2. Constant errors showing control of very few major patterns and frequently preventing communication.
3. Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding.
4. Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding.
5. Few errors, with no patterns of failure.
6. No more than two errors during the interview.

**Vocabulary**

1. Vocabulary inadequate for even the simplest conversation.
2. Vocabulary limited to basic personal and survival areas (time, food, transportation, family, etc.).
3. Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics.
4. Professional vocabulary adequate to discuss special interests; general vocabulary permits discussion of any non-technical subject with some circumlocutions.
5. Professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied social situations.
6. Vocabulary apparently as accurate and extensive as that of an educated native speaker.

**Fluency**

1. Speech is so halting and fragmentary that conversation is virtually impossible.
2. Speech is very slow and uneven except for short or routine sentences.
3. Speech is frequently hesitant and jerky; sentences may be left uncompleted.
4. Speech is occasionally hesitant, with some unevenness caused by rephrasing and groping for words.
5. Speech is effortless and smooth, but perceptively non-native in speed and evenness.
6. Speech on all professional and general topics as effortless and smooth as a native speaker's.

**Comprehension**

1. Understands too little for the simplest type of conversation.
2. Understands only slow, very simple speech on common social and touristic topics; requires constant repetition and rephrasing.
3. Understands careful, somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing.
4. Understands quite well normal educated speech when engaged in a dialogue, but requires occasional repetition or rephrasing.
5. Understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.
6. Understands everything in both formal and colloquial speech to be expected of an educated native speaker.

**WEIGHTING TABLE**

| | 1 | 2 | 3 | 4 | 5 | 6 | (A) |
|---|---|---|---|---|---|---|---|
| Accent | 0 | 1 | 2 | 2 | 3 | 4 | _____ |
| Grammar | 6 | 12 | 18 | 24 | 30 | 36 | _____ |
| Vocabulary | 4 | 8 | 12 | 16 | 20 | 24 | _____ |
| Fluency | 2 | 4 | 6 | 8 | 10 | 12 | _____ |
| Comprehension | 4 | 8 | 12 | 15 | 19 | 23 | _____ |
| | | | | | | Total | _____ |

Note the relative weightings for the various components.

The total of the weighted scores is then looked up in the following table, which converts it into a rating on a scale 0–4+.

**CONVERSION TABLE**

| Score | Rating | Score | Rating | Score | Rating |
|---|---|---|---|---|---|
| 16–25 | 0+ | 43–52 | 2 | 73–82 | 3+ |
| 26–32 | 1 | 53–62 | 2+ | 83–92 | 4 |
| 33–42 | 1+ | 63–72 | 3 | 93–99 | 4+ |

(Adams and Frith 1979: 35–8)

Where analytic scales of this kind are used to the exclusion of holistic scales, the question arises (as with the testing of writing) as to what pattern of scores (for an individual candidate) should be regarded as satisfactory. This is really the same problem (though in a more obvious form) as the failure of individuals to fit holistic descriptions. Once again it is a matter of agreeing, on the basis of experience, what failures to reach the expected standard on particular parameters are acceptable.

The advice on creating rating scales given in the previous chapter is equally relevant here:

### *Calibrate the scale to be used*

Generally the same procedures are followed in calibrating speaking scales as were described for writing scales, with the obvious difference that video-recordings are used rather than pieces of written work.

### *Train scorers (as opposed to interviewers)*

The training of interviewers has already been outlined. Where raters are used to score interviews without acting as interviewers themselves, or are involved in the rating of responses to audio- or video-recorded stimuli, the same methods can be used as for the training of raters of written work.

### *Follow acceptable scoring procedures*

Again, the advice that one would want to offer here is very much the same as has already been given in the previous chapter. Perhaps the only addition to be made is that great care must be taken to ignore personal qualities of the candidates that are irrelevant to an assessment of their

language ability. I remember well the occasion when raters quite seriously underestimated the ability of one young woman who had dyed her hair blonde. In an oral test it can be difficult to separate such features as pleasantness, prettiness, or the cut of someone's dress, from their language ability – but one must try!

## Conclusion

The accurate measurement of oral ability is not easy. It takes considerable time and effort, including training, to obtain valid and reliable results. Nevertheless, where a test is high stakes, or backwash is an important consideration, the investment of such time and effort may be considered necessary. Readers are reminded that the appropriateness of content, of rating scales levels, and of elicitation techniques used in oral testing will depend upon the needs of individual institutions or organisations.

## Reader activities

These activities are best carried out with colleagues.

1. For a group of students that you are familiar with, prepare a holistic rating scale (five bands) appropriate to their range of ability. From your knowledge of the students, place each of them on this scale.
2. Choose three methods of elicitation (for example role play, group discussion, interview). Design a test in which each of these methods is used for five to ten minutes.
3. Administer the test to a sample of the students you first had in mind.
4. Note problems in administration and scoring. How would you avoid them?
5. For each student who takes the test, compare scores on the different tasks. Do different scores represent real differences of ability between tasks? How do the scores compare with your original ratings of the students?

## Further reading

Two books devoted to oral testing and assessment are Luoma (2003) and Underhill (1987). Fulcher (1996a) investigates task design in relation to the group oral. Chahloub-Deville (1995) and Fulcher (1996b) address issues in rating scale construction, the latter with particular

reference to fluency. Kormos (1999) provides evidence that role play can be a useful testing technique, especially when one wants to assess the ability to manage interactions. Lazaraton (1996) examines the kinds of linguistic and interactional support which interlocutors may give to candidates. Douglas (1994) shows how the same rating may be assigned to qualitatively different performances in an oral test. Lumley and McNamara (1995) report on a study into rater bias in oral testing. Wigglesworth (1993) shows how bias in raters can be detected and how raters can improve when their bias is brought to their attention. Shohamy et al. (1986) report on the development of a new national oral test which appears to show desirable psychometric qualities and to have beneficial backwash. Bachman and Savignon (1986) is an early critique of the ACTFL oral interview, to which Lowe (1986) responds. Salaberry (2000) is also critical of it and proposes changes. Shohamy (1994) discusses the validity of direct versus semi-direct oral tests. Powers et al. (1999) report on the validation of the TSE. Luoma (2001) reviews the TSE. The Cambridge CCSE handbook and past papers are a good source of ideas for tasks (address to be found on page 73). Modern 'communicative' textbooks are another source of ideas for tasks. Information on the ARELS examinations (and past papers with recordings) can be obtained from ARELS Examinations Trust, 113 Banbury Road, Oxford, OX2 6JX.

1. Referred to as 'functions' in the handbook.
2. Three tasks are offered for each section but a student only performs one of them. The institution decides which task is most appropriate for each student. As can be seen, only one task for each section is reproduced here.
3. I understand that the FSI no longer tests oral ability in the way that it did. However, I have found the methods described in their 'Testing Kit', which also includes both holistic and analytic scales, very useful when testing the language ability of professional people in various situations.