

Testing oral skills

Bertha Gpe. Zepeda

Lic. Norma Angélica Espinosa Butrón

Information on student achievement is crucial to teaching and learning. As teachers we need information on students' progress; as well as to monitor the functional ability of students, what they can do in relation to the language. This information can be obtained by means of tests. The aim of this paper is to provide information to help pre-service and in-service novice language teachers to write better oral tests; as well as to expose them to a broad repertoire of oral test elicitation techniques.

Oral tests may be considered to be "*the most challenging of all language exams to prepare, administer, and score*" (Madsen, 1983,147), as it is difficult to determine *what* and *how* when it comes to evaluating spoken language. Oral tests are qualitatively different from other kinds of tests; in fact they should be considered as a class of their own.

Underhill (2000) suggests that the following considerations should be taken into consideration when designing an oral test:

- The need to design the test as a whole. An oral test must be conceived as an integral whole. The entire procedure should appear as a single and consistent entity to testers and learners alike.

- The need to make use of a humanistic approach. Oral tests must treat people as human beings. A humanistic approach will possibly allow us to get a much more accurate picture of the test takers oral ability.
- The need to find a suitable balance. The planning and execution of an oral test involves making positive compromises between different forces, for example, between communicative and structural aims, between impression and analytic marking systems, and between highly reliable and highly valid techniques. In addition Hughes (2004) points out that the tasks used to test oral skills –elicitation techniques- should:
 - Represent a sample of the population of oral tasks that we expect candidates to be able to perform.
 - Elicit behavior which truly represents the candidate's ability
 - Be scored validly and reliably.

Framework for testing spoken interaction

Weir (1993) proposes a three-part framework for testing spoken interaction:

1. Covering the operations –tasks/skills- that are involved in spoken interaction such as informal routines, e.g. telling a story, and the improvisational skills that might be called into play when the performance of these routines, breaks down for example: requesting clarification.
2. The conditions under which the tasks are performed for example: time constraints, the number of people involved and their familiarity with each other and
3. The quality of output, the expected level of performance in terms of various relevant criteria e.g. accuracy, fluency or intelligibility.

1. Operations

To test whether learners can speak, it is necessary to get them to take part in direct spoken language activities. We are no longer interested in testing whether

candidates know how to assemble sentences in the abstract: we want candidates to perform relevant language tasks and adapt their speech to the circumstances, making decisions under time pressure, implementing them fluently, and making any necessary adjustments as unexpected problems might arise. The more direct we can make a test and the more contextual and interactional features that can be incorporated of a real-life activity into tests, will reflect as a result a set of statements about what candidates should be able to do in a real-life context. Bygatte (1987) offers a useful description of how speakers organize in routines what they have to communicate. Routines are normally recurring patterns of organization of communication, and can be found firstly in the organization of information, and secondly in the organization of interaction. This author provides an outline of routines and improvisational skills:

Information routines

These routines are conventional ways of presenting information, and are best seen as frequently recurring ways of structuring speech, such as descriptions, comparisons, instructions, telling stories. They can be either expository routines – involve factual information which is typically concerned with questions of sequencing or identity of subject- or evaluative routines –involve drawing conclusions, usually requiring expressions of reasoning, explanations, predictions, justifications preferences and decisions. These are normally made in connection with expository routines. Both types of informational routines can be catered for in test tasks such as oral presentations which cater for long turns, but they may also form part of interactional routines and may be tested through information gap tasks, role plays or interviews.

Interaction routines

This kind of routines can be found in interactions such as buying goods in a shop, or telephone conversations, interviews, meetings, discussions, decision making among others, which on the whole tend to be organized in characteristic ways.

Improvisational skills

If something goes wrong with the information and interactional routines, improvisational skills are called for. Most communication problems which occur in spoken interaction are normally solved through "*improvisational skills*". Bygate (1987) identifies two major kinds of improvisational skills: ***negotiation of meaning and management of interaction***.

- ***Negotiation of meaning***. This is concerned with the way participants contribute to understanding during an interaction. It includes procedures which speakers follow to ensure that understanding takes place. These might include conversational adjustments to maintain contact and knowledge of clarification procedures. Richards et al (1992) conceive negotiation of meaning as what speakers do in order to achieve successful communication. For conversation to progress naturally and for speakers to be able to understand each other it may be necessary for them to:
 1. Indicate that they have understood/do not understand, or that they want the conversation to continue
 2. Help each other to express ideas
 3. Make corrections when necessary to what is said or how it is said, which is commonly defined as: repair.
- ***Management of interaction***. Is related to decisions in terms of who is to speak and what the topic is going to be. Management of interaction covers: Agenda management and turn-taking. Agenda management is concerned with the control over the content, development and duration of the topic; and turn-taking refers to the question of who speaks when and for how long. It may include knowing how to signal that one wants to speak, recognizing the right moment to get a turn, and knowing how to let the other person have a turn.

As a result, test developers must clearly have to specify in the framework of a given test, the type of routines test takers should perform as well as the improvisational skills that might be expected of them.

2. Conditions

Speaking involves the ability to use language to satisfy two particular demands, which we should try and ensure are present in our tests of spoken interaction. The first is that of processing conditions. That is to say, speech takes place under time pressure. It is possible to distinguish between long and short speaking turns. Short turns are more common: usually more spontaneous. Long turns for example, can easily be observed in oral presentations or lectures, which require more planning decisions, and thus often tend to be more prepared.

The second demand is that of appropriate reciprocity conditions. This condition is concerned with the dimension of interpersonal interaction, the relation between speaker and listener. The degree of participation in a developing interaction varies depending on whether it is a lecture, interview, conversation and so on. In some situations a lecture for example, only the speaker normally has speaking rights and takes on almost total responsibility. In a conversation, both participants normally have speaking rights. Clearly if we wish to test spoken interaction, a valid test must include reciprocity conditions. This means that the candidate must be involved in the interaction to a greater extent than that of merely answering questions. Other important aspects that should be taken into consideration are:

- Purpose. An important element to consider amongst the conditions is that of purpose. The purpose of the speakers will help to define the structure and focus of the interaction, Achieving realism in tests for general English is not easy, but the emphasis must still be on giving the interlocutors as realistic, and as needs-based a purpose as possible. Full authenticity of tasks may not be achievable, but we need to make our tests as valid as possible.
- Interlocutors. The number of people involved in the interaction in the test should reflect wherever possible with the situation in real life that one wishes to make statements about. The tester needs to consider with whom the candidates will be using English in their future target situations.

- **Setting.** The setting for the task also needs to be given consideration. Full authenticity of setting is obviously not attainable but should be made as realistic as possible.
- **Channel.** The channel for communication can have an obvious impact on the performance. For example, it may place greater burdens on candidates if they have to simulate a telephone conversation with an interlocutor in a different room as against carrying out a face-to-face conversation in the same room.
- **Input dimensions.** According to Weir (1993) this is perhaps the least definable part of the framework as what is said to the candidate by the other participant(s) will influence his/her own performance. As a result potential sources of variance in performance caused by features of the language used by the interlocutor e.g. rate of utterances, accent of the examiner, clarity of articulation of the examiner, length of discourse –long or short turns- will impact candidate’s performance. The author proposes to give candidates an equal chance to demonstrate their ability; contribution to the interaction should not be widely dissimilar from candidate to candidate. In other words the interlocutor’s contribution to the interaction must be standardized as far as possible.

3. Assessing Quality of Output

The third element that needs to be considered in test task design is how we are to measure the quality of the output which results from the spoken language tasks we adopt. The relationship between a task and the criteria that can be applied to its product is an essential factor on what to include in a test of spoken production. Tasks cannot be considered separately from the criteria that might be applied to the performances they result in. Having established suitable tasks and appropriate assessment criteria to accompany them, considerations need to be given as to how best to apply the criteria to the samples of task performance. There is a need

to establish clear criteria for assessment and to standardize examiners in their use of the proposed criteria.

In oral testing there is a need for explicit comprehensive marking schemes, close moderation of test tasks and marking schemes as well as the training and standardization of markers. In order to measure the quality of spoken performance, we first need to establish criteria of assessment.

Assessing the handling of routine skills:

It is the effectiveness with which appropriate messages are communicated that is normally of interest in testing spoken interaction. To what extent can the candidate demonstrate an ability to meet effectively the demands of required informational or interactional routines? The examiner might make specific assessments in terms of the following:

- Normal time constraints would have to be observed in all required routines. Fluency as overall smoothness of execution of the tasks would be assessed.
- In addition one might want to comment on the discourse coherence, that is, the internal organization of the stages of the discourse. This may be especially relevant in longer turns.
- Appropriateness: this would include the sociocultural ability to take into account setting, topic, role relationships, formality required. Due observance of the norms of interaction in terms of silence, proximity and dealing with encoding difficulties might be looked for.

Assessing the handling of improvisational skills

This might involve the examiner taking a decision on overall effectiveness in two important improvisation abilities:

- a) Ability to negotiate meaning in case of comprehension or production difficulties manifested on the part of the candidate or his/her interlocutor.
- b) Ability to manage interaction (agenda and turn taking) actively and flexibly. This is particularly important where speakers can be expected to be active participants.

If it is necessary to be more specific about the effectiveness in deploying improvisational skills the examiner might make detailed assessments in terms of the following:

- Fluency: that is, smoothness of execution. Ability to negotiate meaning would, for example include the ability to use communication strategies with ease when in difficulties.
- Appropriateness: This could include, for example, the degree of politeness and suitability of timing in turn taking or suitability of the language used in request for clarification or disagreement.

There may also be occasions, for example when the level of the candidates is quite low, that the examiner might need to make assessments at the microlinguistic level.

Assessing the handling of microlinguistic skills

If the focus is on linguistic proficiency at the utterance level, the examiner might wish to use the following criteria:

- Accuracy focusing on both intelligibility and grammar.
- Range: Adequacy and variety of vocabulary employed: adequacy and variety of structures employed.

Appendix 1 shows a set of criteria that might be used to evaluate the various aspects of spoken interaction for assessing the handling of routines and the handling of any necessary improvisation and for operating at the microlinguistic

level. Testers also need to decide whether they will treat these criteria separately in an analytic scheme -lots of separate impressions- or try to collapse them into some form of global impression banding.

As it has been pointed out throughout this paper, the appropriate selection of the elicitation techniques helps to ensure the feature of validity in an oral exam in order to obtain more reliable information with regard to test takers oral performance.

Types of oral test elicitation techniques

There is no natural classification of oral test elicitation techniques, so a list of the most common ones and a brief explanation of each technique will be provided (Underhill: 2000). In some cases, one technique might call on others e.g. skills as well as oral proficiency; a general knowledge of the language or even general knowledge of the world. This is not necessarily a bad thing, and to some extent is unavoidable in tasks which claim to be authentic; such general knowledge contributes to the overall ability to communicate. The broad aim of these techniques is to encourage test takers to speak by giving them something to speak about. The sensitive test designer will always want to adapt ideas to his/her own circumstances to produce the best test:

Discussion/conversation technique

This is the most natural thing in the world, two people having a conversation on a topic of common interest. It is also the hardest to make happen in the framework of a language test; it can only occur when both parties are relaxed and confident and something sparks between them, allowing the activity -a conversation- to become dominant, and its ulterior purpose -a language test- to be temporarily subordinated. Taking the initiative, asking questions, expressing disagreement, all require a command of a particular language feature.

Oral report technique

The learner prepares and gives an oral presentation lasting from five to ten minutes. He/she is expected to refer to notes, but reading aloud is strongly discouraged. The use of simple aids such as an overhead projector, blackboard or flipchart diagrams is encouraged if appropriate. At the end of the presentation, the speaker is expected to deal with any questions.

Variation 1: identifying a topic of personal interest

Instead of imposing a limited choice and formal preparation time on the learner, another way of eliciting a short presentation is to try consciously to identify possible topics in earlier stages of the test. For example, the oral test can begin with a question and answer routine. Some examples of such probing questions are:

- What do you do in your spare time?
- What kinds of books do you like reading?
- What did you do on your last holiday?
- Do you enjoy traveling? Where to? Why?

Test takers joint discussion/decision making technique

A group of two or more learners are tested together, without the participation of an interviewer. The learners have to maintain and direct the discussion entirely on their own. The task usually involves taking information from written documents and coming to a decision or consensus about certain questions through group discussion. Where several documents or sources are used, these can be read before the discussion begins.

Role play technique

The learner is asked to take on a particular role and to imagine himself in that role in a particular situation. He has to converse with the interviewer in a way that is appropriate to the role and the situation given. The learner is given a set of instructions, just before the test, that explain in simple language exactly what he is supposed to do. For more confident learners, these instructions may be expressed in terms of the general situation; or the instructions may be made more specific.

The ability to ask questions is important in this particular technique. This skill is often overlooked in oral tests and it is something role-plays are good at eliciting. Role-play situations may be chosen to test the learner's command of general social language, or to elicit particular types of language.

Interview technique

The interview is the most common of all oral tests; for many people it is the only kind of oral tests. It is a direct, face to face exchange between learner and interviewer. It follows a pre- determined structure, but still allows both people a degree of freedom to say what they genuinely think. Interviewers usually have a prepared list of written or memorized questions to ask, or topics to bring up. This mental or written list will contain quite a wide variety of questions and topics in order to avoid constant repetition and possible compromise. An interview is more authentic; it has a consistency and a relevance that stretches over more than one question or comment. There may be several topics raised in an interview, but each is explored in enough detail, with follow- up questions and prompting, to allow the learner to develop it and to show his proficiency, rather than just giving a straight answer to a straight question.

Form –filling technique

The test taker and interviewer work together to fill in a form. The questions usually concern the test taker's personal details, professional situation or language needs. Either the interviewer or the test taker may actually write in the answer; if the test taker does it, the test will take longer and the test taker will obviously feel his writing as well as his/her speaking skills are being tested. Alternatively, the test taker can be given ten to fifteen minutes to fill in the form before the oral discussion; in this case the form-filling serves to elicit information that the interviewer can use subsequently for discussion.

Variation 1 using a questionnaire

This is a widely-used testing technique. Compared to a form, which usually asks for factual personal details or history, a questionnaire asks about one specific area of personal tastes or preferences in some detail such as food, drink, music, entertainment, TV programmes, sport, hobbies, and holidays, among others.

Using a picture and picture story technique

The test taker is given a picture or sequence of pictures to look at. Then the interviewer asks the learner to describe the picture or story and allows him to speak freely. When the learner has finished speaking, the interviewer may ask questions designed to elicit particular information, perhaps about a point the test taker has missed or not made clear. As soon as the interviewer begins to ask questions, the test taker will tend to assume that he is now only expected to answer questions and is no longer expected to continue his commentary. Usually, the interviewer asks general questions intended to elicit the test taker attitudes or opinions on a topic related to the subject of the picture, but not directly about it. This is a common way of leading into a discussion.

A picture or cartoon story usually consists of from four to twelve drawings telling a story which is simple but allows the test taker to add his own interpretation about the people or events involved. The pictures are specially drawn for this purpose, to exclude objects or actions which are difficult to describe or are very cultural driven. The pictures can be exploited at two levels: description of the people, objects and events, leading into an interpretation and discussion of the more general issues that have been raised. Visual stimuli are an economic and effective way of providing a topic of conversation without giving the test taker words and phrases to manipulate and give back.

Variation 1: using several similar pictures

Another alternative is to use “*spot-the difference*” type cartoons; the test taker describes the differences between the pictures. The important point of this kind of task is to assess the fluency and facility of the language generated, and not the learner’s visual acuity.

Giving instructions/ descriptions/ explanation technique

With minimal preparation, the test taker describes, at some length, a well known object, a system or an everyday procedure. The description is factual and the object being described either widely known or easily comprehensible. Choosing something that is familiar to everybody is a good way of getting the test taker to produce connected discourse on a given topic but allows considerable freedom of choice of expression without requiring extensive preparation. This distinguishes it from the oral report, which requires more thorough preparation on a more specialized subject. Normally, the test taker is given a list of between five to seven topics to choose from and a few minutes’ preparation time. Some examples of suitable topics are:

- How do you make a good cup of tea or coffee?

- Give instructions for using a public pay- phone.
- Explain how you would advise someone to look for a job.

The choice of topics can make the task more or less controlled. A question such as *'Describe your favorite meal'* would be less controlled as there can be a lot of possible answers; whereas *"Give instructions for using a public pay-phone"* has basically only a single answer -in content, but many different ways of phrasing it-. The important feature of all these topics is that in order to be satisfactorily carried out they require the test taker to express at some length say, a minimum of six to eight sentences.

The interviewer should be aware of the possibility of memorized speeches being worked into topics to which they are marginally relevant. If these memorized segments are large- two or three sentences long- they should be noticeable as being either stylistically unusual or irrelevant to the topic. The instructions and the marking system should make it clear that marks are awarded only for what is relevant to the subject. On the other hand, if the memorized segments are short, i.e. phrase length, they will probably be unnoticeable, and should arguably be rewarded anyway if they increase the overall fluency

To sum up, it should also be pointed out that in order to make an oral test valid and reliable a representative sample of the oral tasks that the candidates are expected to be able to perform should be set. This means that a single oral test elicitation technique should not be provided, more than one task is required. In addition Hughes (2004) suggests to plan and structure the test carefully by taking into consideration the following recommendations:

- Make the oral test as long as feasible.
- Plan the test carefully
- Give the candidate as many "fresh starts" as possible
- Use a second tester for interviews
- Collect enough relevant information

- Carry out the interview in a quiet room with good acoustics
- 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.
- Select interviewers carefully and train them.

References

- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.
- Madsen, H.S. (1983). *Techniques in Testing*. Oxford: Oxford University Press.
- Richards, J.C., Platt, H. (1992). *Longman Dictionary of Language Teaching and Applied Linguistics*. England: Longman Group UK.
- Underhill, N. (2000). *Testing Spoken Language*. Cambridge: Cambridge University Press.
- Weir, C.J. (1993). *Understanding and Developing Language Tests*. Hemel Hempstead: Prentice Hall

Appendix 1 Analytic marking scheme

Criteria of assessment
<p>Appropriateness</p> <p>0. Unable to function in the spoken language.</p> <p>1. Able to operate only in a very limited capacity: responses characterized by sociocultural</p> <p>Inappropriateness.</p> <p>2. Signs of developing attempts at response to role, setting, etc., but misunderstanding may occasionally arise through inappropriateness, particularly of sociocultural convention.</p>

3. Almost no errors in the sociocultural conventions of language; errors not significant enough to be likely to cause social misunderstandings.

Adequacy of vocabulary for purpose

0 Vocabulary inadequate even for the most basic parts of the intended communication.

1 Vocabulary limited to that necessary to express simple elementary needs; inadequacy of vocabulary restricts topics of interaction to the most basic: perhaps frequent lexical inaccuracies and/or excessive repetition.

2 Some misunderstandings may arise through lexical inadequacy or inaccuracy; hesitation and circumlocution are frequent, though there are signs of a developing active vocabulary.

3 Almost no inadequacies or inaccuracies in vocabulary for the task. Only rare circumlocution

Grammatical Accuracy

0 Unable to function in the spoken language; almost all grammatical patterns inaccurate, except for a few stock phrases.

1 Syntax is fragmented and there are frequent grammatical inaccuracies; some patterns may be mastered but speech may be characterized by a telegraphic style and/or confusion of structural elements.

2 Some grammatical inaccuracies: developing a control of major patterns, but sometimes unable to sustain coherence in longer utterances.

3 Almost no grammatical inaccuracies; occasional imperfect control of a few patterns.

Intelligibility

0 Severe and constant rhythm, intonation and pronunciation problems cause almost complete unintelligibility.

- 1 Strong interference from L1 in rhythm, intonation and pronunciation; understanding is difficult, and achieved often only after repetition.
- 2 Rhythm, intonation and pronunciation require concentrated listening, but only occasional misunderstanding is caused or repetition required.
- 3 Articulation is reasonably comprehensible to native speakers; there may be a marked 'foreign accent' but almost no misunderstanding is caused and repetition required only infrequently.

Fluency

0 Utterances halting, fragmentary and incoherent.

1 Utterances hesitant and often incomplete except in a few stock remarks and responses. Sentences are, for the most part, disjointed and restricted in length.

2 Signs of developing attempts at using cohesive devices, especially conjunctions. Utterances may still be hesitant, but are gaining in coherence, speed and length.

3 Utterances, whilst occasionally hesitant, are characterized by an evenness and flow hindered, very occasionally, by groping, rephrasing and circumlocutions. Inter-sentential connectors are used effectively as fillers.

Relevance and adequacy of context

0 Response irrelevant to the task set; totally inadequate response.

1 Response of limited relevance to the task set; possibly major gaps and/ or pointless repetition.

2 Response for the most part relevant to the task set, though there may be some gaps or redundancy.

3 Relevant and adequate response to the task set.

Source: TEEP, CALS, University of reading.